The Dissertation Committee for Ryan Cheng
certifies that this is the approved version of the following dissertation:

Conformational Dynamics of an Unfolded Biopolymer: Theory and Simulation

Committee:

Dmitrii E. Makarov, Supervisor

Ron Elber

Adrian T. Keatinge-Clay

Graeme Henkelman

Ernst-Ludwig Florin

Conformational Dynamics of an Unfolded Biopolymer: Theory and Simulation

by

Ryan Cheng, B.S.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

December 2012

Dedicated to my parents

Acknowledgments

First and foremost, I would like to thank my advisor Professor Dmitrii Makarov for his kindness, patience, and vast knowledge. It has been a pleasure working with Dima and I have spent much of the last five years trying to learn as much as possible from such an outstanding scientist and teacher. I am also grateful for my labmates Sai Konda and Alex Hawk, who are both thoughtful and filled with insight, making our office a stimulating and engaging place to work (and also for playing "office pong" with me once in a while). I would like to thank Gizelle Sherwood and Professor Linda Peteanu for introducing me to chemical research many moons ago and Lei Huang, Sylvia Jean, Penny Kile, Serdal Kirmizialtin, and Betsy Hamblen for offering me their great advice from time to time. Furthermore, I am fortunate to have collaborated with a number of great scientists during my time in graduate school: Takanori Uzawa, Professor Kevin Plaxco, Craig Cone, Professor David Vanden Bout, James Madsen, Professor Jennifer Brodbelt, Andrea Soranno, Professor Ben Schuler, and Alex Hawk.

I would like to thank my parents for their love, support and encouragement during my time in graduate school (and the great people at Skype for creating an effective medium through which I could communicate regularly with my parents). I would also like to thank my brilliant sisters whom I have

looked up to as long as I could remember.

Including people I have already mentioned, I would also like to thank some of the people I've been lucky enough to meet in Austin: Taki Adachi, Amy Bonaparte, Katie Clark, Craig Cone & Courtney Russo, Michelle Fox, Micah Glaz, Zach Pozun, Rye Terrell, and Ting Yan. Thank you for your friendship and kindness, the million times we ate at Chuy's, the countless LOLs, and the many cherished memories.

Conformational Dynamics of an Unfolded Biopolymer: Theory and Simulation

Ryan Cheng, Ph.D.
The University of Texas at Austin, 2012

Supervisor: Dmitrii E. Makarov

The conformational dynamics of an unfolded biopolymer such as a polypeptide or DNA has attracted a significant amount of attention in the context of protein folding and the design of biomimetic technologies. To this end, recent advances in single-molecule experiments have allowed for biomolecules to be probed with an unprecedented level of detail, shedding light on their dynamics. Motivated by the need to interpret experimental data and to help guide future studies, we use concepts from polymer physics, computer simulations, and experimental data to study the timescales in which an unfolded biopolymer undergoes conformational rearrangement.

First, we examine the end-to-end loop formation time in the experimentally relevant scenario where the dynamics are probed using a fluorescence probe and quencher. We show that the loop formation time in the experimentally relevant case is quantitatively dissimilar from the predictions of previous

theoretical studies that neglect the quenching kinetics, which are often used to interpret experimental data.

We additionally find that the loop formation times can be re-casted in a simple, universal dependence that is characteristic of random-coils. Furthermore, deviations from this universal dependence can be used as a sensitive tool for detecting structural order in unfolded biopolymers.

We also consider a surface-tethered polymer chain and investigate the rate of a reaction between the free end and the surface. We explore this rate in the reaction-controlled limit and the diffusion-controlled limit, providing evidence for near-universal dependences of the rate in the respective limits.

Next, we examine the transit time of end-to-end loop formation in a case study. We find that approximating the end-to-end dynamics as diffusion in a 1D potential of mean force fails dramatically to describe the transit time. Furthermore, we find that the transit time is uninfluenced by the average entropic force imposed by the polymer chain and is well described by a simple free-diffusion model.

Finally, we explore the role of internal friction in the dynamics of an unfolded protein. Using simple polymer models that incorporate internal friction as an adjustable free parameter, we mimic typical single-molecule experiments that probe the unfolded state dynamics and make several experimentally verifiable predictions.

Table of Contents

List of Figures

Chapter 1

Introduction

A central theme of the work described in this dissertation is the use of simple polymer models and concepts from polymer physics to gain insight on the dynamics of unfolded proteins and single-stranded oligonucleotides. The work described herein uses a combination of theory and computer simulation, often in conjunction with experimental data from collaborators or from published works. The proceeding sections will discuss some of the motivations for this work and provide a summary of the proceeding chapters, respectively.

1.1 Motivations

Understanding the timescales in which an unfolded biopolymer undergoes conformational rearrangement is of great interest in the scientific community in a number of contexts. In particular, a significant amount of focus has been directed to understanding how an unfolded protein, which exhibits a multitude of unfolded state conformations, folds into a native conformation (Fig. 1.1). To this end, the timescales in which an unfolded protein interconverts between nonnative conformations is intimately related to the "speed limit" of protein folding [1]. Furthermore, a number of studies have focused

on polymer rearrangements that mimic elementary steps in folding such as the process in which two distant residues of a polypeptide diffusively come into contact with one another to form a loop, which is thought to play a key role in protein folding [2–5].

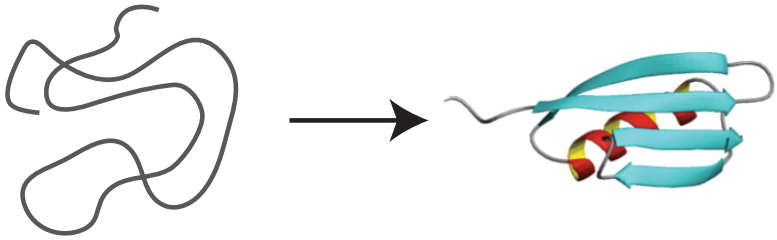


Figure 1.1: Cartoon of protein folding from a disordered, random-coil conformation to the native conformation

Likewise, the dynamics of oligonucleotides such as single-stranded DNA (ssDNA) have attracted a significant amount of attention due to its importance in biological processes such as gene expression and regulation (Ref. [6] and Refs. therein). Furthermore, ssDNA has also played an increasingly important role in the design of biomimetic technologies, such as the electrochemical DNA (E-DNA) sensor [7, 8], which utilizes a change in the conformational dynamics of DNA upon the binding of a target molecule. In the simplest description, this sensor consists of single-stranded DNAs that are attached at one end to an electrode surface while the free end is attached to a redox group (Fig. 1.2). In the absence of the complementary target sequence, the surface-tethered DNAs remain flexible such that the redox-groups can readily approach the surface,

resulting in a detectable signal. However, when the target sequence is present, there is a significant drop-off in the detectable signal due to the formation of rigid double-stranded DNAs that inhibit the redox groups from approaching the surface.

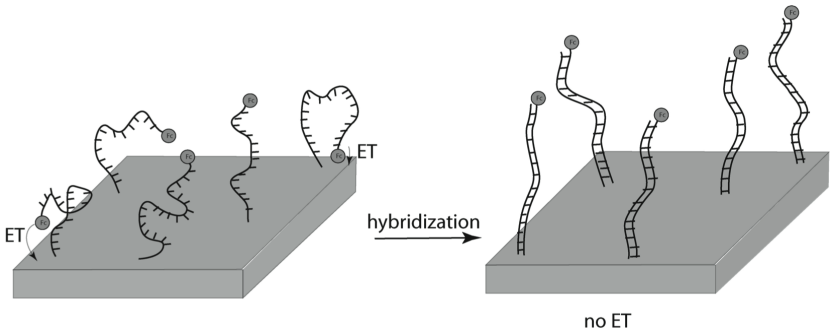


Figure 1.2: Electrochemical DNA (E-DNA) sensor (Adapted from Ref. [7]). Binding of the surface-attached DNAs to the complementary target molecules inhibits electron transfer between the terminal redox group and the electrode surface.

Thus, understanding the dynamics of unfolded biopolymers is important to the study of problems of biological relevance such as protein folding as well as the design of new biomimetic technologies. To this end, single-molecule fluorescence experiments have been used to study unfolded biopolymers. In one such experiment, a fluorescence probe is attached to one monomer of the polymer chain while a quencher group is attached to a distant monomer. By using a laser pulse to excite the probe and measuring its excited state lifetime, an experimentalist can determine the timescale in which the probe and quencher diffusively come into contact with one another (i.e., forming a loop), thereby quenching the probe (Fig. 1.3). A second type of exper-

iment uses single-molecule nanosecond fluorescence correlation spectroscopy (nsFCS), which employs fluorescence resonance energy transfer (FRET) between a pair of dyes attached to distant monomers along a polymer chain (Fig. 1.4). By monitoring the fluctuations in the fluorescence intensities of the donor and acceptor, which are intimately related to the fluctuations in the distance between the donor and acceptor, one can obtain a characteristic timescale for these fluctuations [9–11].

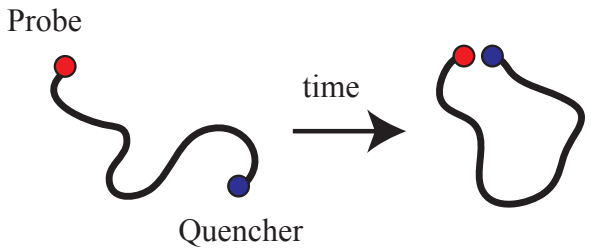


Figure 1.3: Cartoon of loop formation experiment where the fluorescence probe and the quencher group are attached at opposite ends of the polymer chain.

Motivated by these single-molecule experimental studies of unfolded biopolymers, an important goal of the work described in this dissertation is to help interpret data from these experiments and to guide future studies. Furthermore, we are motivated to help aid in the design of more efficient biomimetic technologies like the E-DNA sensor. To address these goals, we often turn to simple polymer models and the use of computer simulations, mimicking the typical single-molecule experiments that probe the dynamics of unfolded biopolymers. Furthermore, our findings are often corroborated by

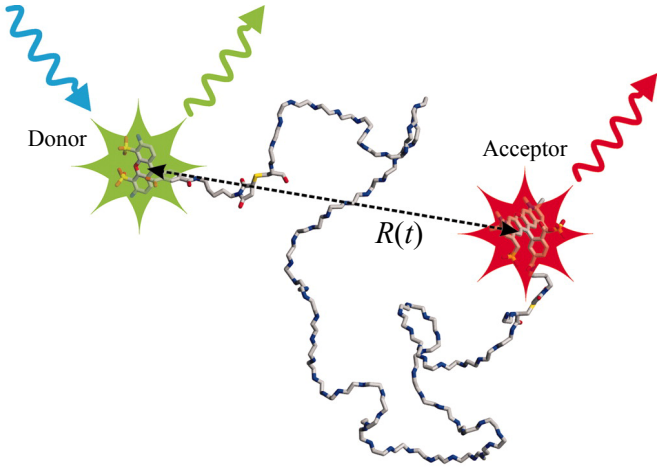


Figure 1.4: Fluctuations in the distance between the donor and acceptor, $R(t)$, are monitored using FRET (Adapted from Ref. [10]).

experimental data either from our close collaborators or from literature.

1.2 Overview of Dissertation

Each of the proceeding chapters are self-contained stories. Their contents are as follows:

In Chapter 2, we explore the rate of end-to-end loop formation of a polymer chain in the experimentally relevant scenario where the rate is measured by monitoring the excited state lifetime of a fluorophore attached to one end of the polymer that is quenched by a quencher group attached to the opposite end. Previous theoretical studies have typically considered the rate at which the respective ends of the polymer come within a cutoff distance from one another without considering the kinetics of the probe/quencher pair.

Here, we mimic an experimental probe and quencher by modeling a quenching process that depends exponentially on the distance between the polymer ends. Using brute force simulations of a bead-and-spring polymer and theoretical arguments, we find that the effective end-to-end loop formation rate in this experimentally relevant case is quantitatively dissimilar from previous theoretical predictions, which are often used to interpret experimental data. Furthermore, our work is corroborated by experimental loop formation data for single-stranded DNA provided by Dr. Takanori Uzawa and Professor Kevin Plaxco. The work described in Chapter 2 was published in the Journal of Physical Chemistry B in 2009 [12].

In Chapter 3, we explore the dynamics of a single polymer chain where one end is fixed on an impenetrable surface (i.e., a surface-attached polymer). Motivated by the E-DNA sensor [7, 8], we examine the rate of a reaction between the free end of the surface-attached polymer and the surface. We consider this rate under two limiting cases: the diffusion-controlled limit, where the diffusion of the free end to within a certain distance of the surface is rate limiting, and the reaction-controlled limit, where the reaction between the free end and the surface is rate limiting. We explore the end-to-surface reaction rate using brute force simulations of a bead-and-spring polymer model and using the Szabo-Schulten-Schulten approximation [13] for a surface-attached Gaussian chain. Furthermore, we compare out results with the better studied end-to-end reaction within an unconstrained polymer chain (i.e., end-to-end loop formation). This work was published in the Journal of Physical Chemistry

B in 2010 [14].

In Chapter 4, we consider the problem of loop formation not only between the ends of a polymer chain, but also between an end monomer and a monomer located at an internal position or between two internal monomers. We explore the kinetics of loop formation using a combination of theory, simulation, and experimental data in collaboration with Dr. Takanori Uzawa and Professor Kevin Plaxco. We address several questions including how the loop formation time is affected by decreasing the loop length while keeping the total chain length fixed. An example of this scenario is if a probe is attached at one end of the polymer chain while the position of the quencher group is systematically moved closer and closer towards the probe. Furthermore, we address whether the loop formation time exhibits generalizable behavior that not only applies to a simulated bead-and-spring homopolymer but also to real polymers such as single-stranded DNAs and unfolded proteins. Chapter 4 was published in the Biophysical Journal in 2010 [15].

In Chapter 5, using end-to-end loop formation in a case study, we examine the duration of a loop formation transition using computer simulations of a bead-and-spring homopolymer. The duration of a transition event (or transit time) refers to the time spent in transition from a subset of conformations A to a subset of conformations B without first returning to A. In the case of loop formation, we examine the transit time for a polymer chain starting from the subset of conformations where the end-to-end distance is r_A and ending with an end-to-end distance of r_B where $r_A > r_B$. Furthermore, we compare

our simulation results to an approximation of the transit time that assumes that the end-to-end dynamics of the polymer is described by 1D diffusion in a potential of mean force. Our study points out several interested properties of the transit time and cautions the use of the aforementioned approximation. Chapter 5 was published in the Journal of Chemical Physics in 2011 [16].

Finally, in Chapter 6, we explore the role of non-solvent related viscosity (i.e., internal friction) in typical single molecule experiments that probe the dynamics of unfolded proteins. We use the Rouse with internal friction (RIF) model as the framework for our study because it incorporates internal friction as a free adjustable parameter. We also introduce the non-free draining form of RIF, Zimm with internal friction (ZIF), which was derived by Alex Hawk. Using these simple models, we mimic typical single molecule experimental studies that probe the dynamics of unfolded biopolymers: reconfiguration dynamics (i.e., FRET) and loop formation dynamics. Our study makes several experimentally testable predictions on the role of internal friction in the unfolded state dynamics of a protein, which warrant future studies. This work has yet to be published.

Chapter 2

The rate of end-to-end loop formation between a fluorescence probe and a quencher group that are attached to opposite ends of an unfolded biopolymer

2.1 Abstract

The problem of determining the rate of end-to-end collisions for polymer chains has attracted the attention of theorists and experimentalists for more than three decades. The typical theoretical approach to this problem has focused on the case where a collision is defined as any instantaneous fluctuation that brings the chain ends to within a specific capture distance. In this paper, we study the more experimentally relevant case, where the end-to-end collision dynamics are probed by measuring the excited state lifetime of a fluorophore (or other lumiphore) attached to one chain end and quenched by a quencher group attached to the other end. Under this regime, a "contact" is defined not by the chain ends approach to within some sharp cutoff but, instead, typically by an exponentially distance-dependent process. Previous theoretical models predict that, if quenching is sufficiently rapid, a diffusion-controlled limit is attained, where such measurements report on the probe-independent, intrinsic end-to-end collision rate. In contrast, our theoretical

considerations, simulations, and an analysis of experimental measurements of loop closure rates in single-stranded DNA molecules all indicate that no such limit exists, and that the measured effective collision rate has a nontrivial, fractional power-law dependence on both the intrinsic quenching rate of the fluorophore and the solvent viscosity. We propose a simple scaling formula describing the effective loop closure rate and its dependence on the viscosity, chain length, and properties of the probes. Previous theoretical results are limiting cases of this more general formula.

2.2 Introduction

The two ends of a polymer chain that diffuses freely in solution are occasionally found in close proximity. The problem of determining the time scale over which such collisions take place has received considerable attention over the last three decades [17–31]. This problem arises in a number of different contexts. In particular, diffusional search for certain native contacts has been proposed to be the rate limiting step in protein folding [3, 5, 32, 33], thus suggesting that the loop closure rate of polypeptides ultimately determines the folding "speed limit" [2, 4]. Thus motivated, this problem has received the attention of several experimental groups, who have measured loop closure times for a range of unstructured polypeptides [2,22,34,35] and single-stranded DNAs [6, 36, 37].

Despite its apparent simplicity, calculation of polymer loop closure times remains an open theoretical problem. The two most commonly used ap-

proximations, the Szabo-Schulten-Schulten (SSS) theory [13] and the Wilemski-Fixman (WF) approximation [31], utilize two different local equilibrium approximations for the polymer dynamics and cannot be systematically improved in any straightforward way. Computer simulations [18, 19, 22–25, 28] have provided further insights and revealed the limitations inherent in these approximations. A recent paper by Thirumalai and co-workers [28] presents a comprehensive comparison of various theories with the results of such simulations.

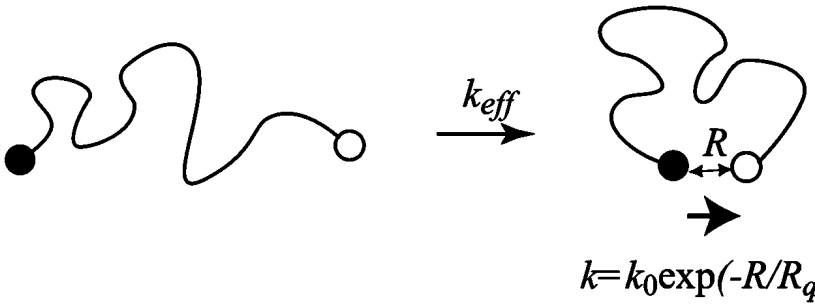


Figure 2.1: The effective end-to-end collision rate is typically measured by monitoring the quenching of a fluorescent probe (or other lumiphore) upon its collision with a quencher attached to the other end of the polymer chain. For electron-transfer-based quenching, which is commonly employed experimentally, the efficiency of the quenching mechanism depends exponentially on the end-to-end distance.

In addition to the above-mentioned approximations inherent in existing theories of polymer loop closure, there remains an additional, fundamental inconsistency between theoretical models of the process and the experimental measurements commonly used to test them. Experimental studies of polymer loop closure rates generally employ probes attached to the chain termini (Fig. 2.1). These typically consist of an optically excited probe (a fluorophore

or other lumiphore) at one end that is quenched when it approaches a quenching group on the opposite terminus. The rate of quenching $k = k(R)$ depends on the distance R between the ends. For example, when the quenching mechanism involves electron transfer, this dependence is given by an exponential function

$$k(R) = k_0 \exp(-R/R_q) \tag{2.1}$$

where k_0 is a parameter describing the intrinsic quenching rate and R_q sets a length scale over which quenching can take place. Typically, this length is on the order of 1 Å. In the following, we will refer to Eq. 2.1 as the "exponential quenching model" (EQM). In contrast to the experimental approach, theoretical studies and computer simulations commonly assume an idealized situation where a collision between the polymer ends takes place when R falls below a certain capture radius. This corresponds to

$$k(R) = k_0 \theta(R_q - R) \tag{2.2}$$

where $\theta(x)$ is the Heaviside step function and $k_0 \to \infty$. We will refer to this model as the contact quenching model (CQM) and will consider the more general situation of finite k_0. In CQM, the overall rate becomes independent of k_0 when k_0 is large enough: the quenching itself is effectively instantaneous as the distance R_q is reached. Therefore, the overall rate is determined by the diffusive dynamics of the polymer ends rather than by the kinetics of quenching.

Here, we show that the difference between the idealized CQM and more

realistic experimental scenarios such as EQM is not merely quantitative. In the latter case, the time scales of the probe quenching cannot be decoupled from those of polymer dynamics. As a result, although such a limit is well-defined for the idealized CQM case, the commonly assumed diffusion-controlled limit, in which the measured quenching rate becomes independent of k_0, does not in fact exist in any experimentally relevant situation. By contrast, the standard SSS and WF approximations, which are widely invoked to interpret experimental data, cannot differentiate between the two cases: They correctly predict the existence of a diffusion-controlled limit for CQM but erroneously predict the same limit for the more realistic EQM. We note that these findings are independent of the exponential form assumed for the distance-dependent quenching in EQM and are valid for any $k(R)$ that does not identically vanish at any finite distance (as in Eq. 2.2).

The finding that the WF/SSS approximations break down at large values of k_0 is not new. In fact, Wilemski and Fixman anticipated such a possibility in their original work [31]. Numerical studies by Srinivas et al. [27] and by Barzykin et al. [17] report significant deviations from the WF result at large values of k_0 for the extreme (and somewhat obvious) experimental case in which $k(R)$ describes the relatively long-range R^{-6} distance dependence of fluorescence resonance energy transfer (FRET). Specifically, in the FRET case, the spatial range of $k(R)$ is often comparable with the polymer length scales, leading to significant and obvious competition between energy transfer and polymer dynamics. Here, however, we address this same issue

for cases in which the end-to-end "collisions", as reported by the quenching rate $k(R)$, are well localized spatially (i.e., for which the quenching mechanism exhibits a more realistically strong distance dependence). Under these conditions, which are true of most of the reported experimental studies of loop closure, the quenching length scale R_q is much shorter than the typical length scale for the polymer's end-to-end distance. The lack of a diffusion-controlled limit and the failure of the standard approximations is far less obvious in this case. Nevertheless, we will show that replacing the commonly employed, finite range transfer model (Eq. 2.2) by a model capturing a more realistic distance dependence (Eq. 2.1) leads to a qualitatively different dependence of the observed collision rate on the properties of the quencher, the polymer, and the solvent.

Our findings make it imperative to reconcile previous experimental data, which were analyzed in terms of WF or SSS theories and/or by using CQM, with the present theoretical predictions. If the diffusion-controlled limit is not achieved experimentally, is it meaningful at all to fit experimental data to approximate theories that predict such a limit? Likewise, since most of the computer simulations reported to date were based on the assumption of hard-shell contact quenching [23, 24, 28], where the diffusion-controlled limit is well-defined, do they have any bearing on the results of experimental loop closure studies? To answer these questions, it is useful to consider our results from a different perspective. The value of k_0 is normally not an experimentally adjustable parameter. Instead, the existence of the diffusion-controlled

rate is commonly ascertained by examining the dependence of the measured effective end-to-end collision rate, k_{eff}, on the solvent viscosity, η. In the diffusion-controlled limit, one expects $k_{eff} \propto \eta^{-1}$. More generally, WF and SSS theories predict that the rate is given by

$$k_{eff}^{-1} = k_D^{-1} + k_R^{-1} \tag{2.3}$$

where $k_D \propto \eta^{-1}$ is the diffusion-controlled rate and k_R is the reaction-controlled rate that is proportional to k_0 and is independent of the viscosity. This suggests that the effective collision time k_{eff}^{-1} should depend linearly on η, while the intercept of the k_{eff}^{-1} vs. η plot yields the reaction-controlled rate [22]. We will show that, for CQM, this equation adequately captures the viscosity dependence of the rate. That is, as the viscosity (or k_0) is increased, one continuously goes from the reaction-controlled regime (where k_{eff} is proportional to the intrinsic quenching rate k_0 and is independent of viscosity) to the diffusion-controlled limit, where k_{eff} is independent of k_0 and is inversely proportional to the viscosity.

We will further demonstrate that the viscosity (and k_0) dependence of the overall rate is more complicated when the more realistic EQM is considered. As the viscosity (or k_0) is increased, three regimes are encountered. In regime I, the WF approximation holds, as described by Eq. 2.3. Note that regime I should not be used synonymously with a diffusion-limited regime: As will be seen below, the system commonly leaves regime I and enters regime II before the diffusion-controlled limit is attained. In regime II, which thus replaces the

familiar diffusion-limited regime expected for CQM, the effective end-to- end collision rate exhibits a power-law dependence on η and k_0 which is of the form

$$k_{eff} \propto k_0^{\delta} \eta^{-(1-\delta)} \tag{2.4}$$

Note that the scaling exponents for k_0 and η are related. We will show that this is necessarily the case as long as the polymer dynamics are overdamped. Finally, in regime III, the solvent is effectively frozen, the quenching kinetics are no longer exponential, and the overall quenching time scales as k_0^{-1} and is independent of viscosity.

The scaling exponent δ observed in regime II is central to the present study. It is not universal but rather depends on the properties of both the polymer and the quenchers. We will show that it generally becomes smaller when the quenching length scale R_q becomes shorter. When δ is close to 0, it becomes difficult to distinguish between the fractional power law of Eq. 2.4 and the η^{-1} dependence expected in the diffusion-controlled limit. Thus, fitting the experimental data with Eq. 2.3 can appear adequate even when the interpretation of the data in terms of the simple CQM is not, a cautionary note that is a key result of the work presented here.

Although simulations can be used to estimate the value of the scaling exponent, they are not always practical. Thus, it is desirable to have a simple estimate of δ based on experimental parameters. In the following, we will present evidence that δ exhibits a universal dependence on the ratio of the polymer's quadratic mean end-to-end distance and the quenching length R_q.

This enables us to estimate the values of δ for various experiments and to reanalyze the experimental data from the point of view of Eq. 2.4.

For a theorist, our study highlights the observation that efforts to generate more realistic theory/simulations may lead to additional complications. If, instead of the idealized CQM, one adopts a more realistic spatial dependence of the quenching rate, then additional information regarding the time scale of the intrinsic quenching rate is required. In contrast to the case with CQM, in this case, the limit of an infinitely fast quenching rate produces a physically meaningless result.

The rest of this paper is organized as follows. In Section 2.3, we present our theoretical arguments and simulation results and explore the dependence of the loop closure rates on the solvent viscosity, and on the spatial range and the intrinsic quenching rate of the probe. In Section 2.4, we compare our predictions with several experimental studies. Finally, Section 2.5 summarizes our main findings, outlines possible extensions, and discusses implications of our results for experimental loop closure studies.

2.3 Results from theory and simulation

2.3.1 Absence of a diffusion-controlled limit

To demonstrate that a breakdown of diffusion-limited kinetics is mathematically inevitable (except for very special cases), consider first the extreme limit of infinite intrinsic quenching rate, $k_0 \to \infty$. Suppose we start with an equilibrium ensemble of terminally labeled polymer chains free in solution. A

laser pulse excites the fluorophores so that at $t = 0$ we have an ensemble of electronically excited molecules with an equilibrium distribution $p(R)$ for the end-to-end distance R (see Fig. 2.2a). To measure the overall quenching rate k_{eff}, we monitor the population of excited molecules, which is related to the survival probability $S(t)$ that a fluorophore is still excited at time t.

Let us first consider the CQM case (Eq. 2.2), in which quenching occurs whenever the chain ends happen to be within a capture distance R_q. In the $k_0 \to \infty$ limit, the molecules with $R < R_q$ will be quenched instantly (dark shaded area in Fig. 2.2a). Thus, $S(t)$ will exhibit a sudden drop from 1 to $\int_0^{R_q} p(R)\,\mathrm{d}R$ (see Fig. 2.2b). The population of excited molecules at $t = 0+$ will thus have a "hole" at short distances (as described by the lightly shaded area in Fig. 2.2a). If diffusional rearrangement of the chains is neglected, the molecules with fluorophores outside this "hole" will remain excited indefinitely (in our idealized model) because their rate of quenching, according to Eq. 2.2, is identically zero. Thus, $S(t)$ will remain constant (as shown by a dashed line in Fig. 2.2b). When polymer diffusion is present, however, the fluorophores originally outside of this hole will eventually approach to within $R < R_q$, resulting in a decreasing survival probability. It is clear from this argument that the measured overall quenching rate is diffusion limited. If the survival probability decays exponentially

$$S(t) \propto \exp(-k_D t) \tag{2.5}$$

then Eq. 2.5 defines the diffusion-controlled rate constant k_D, which vanishes in the limit of infinitely slow diffusion.

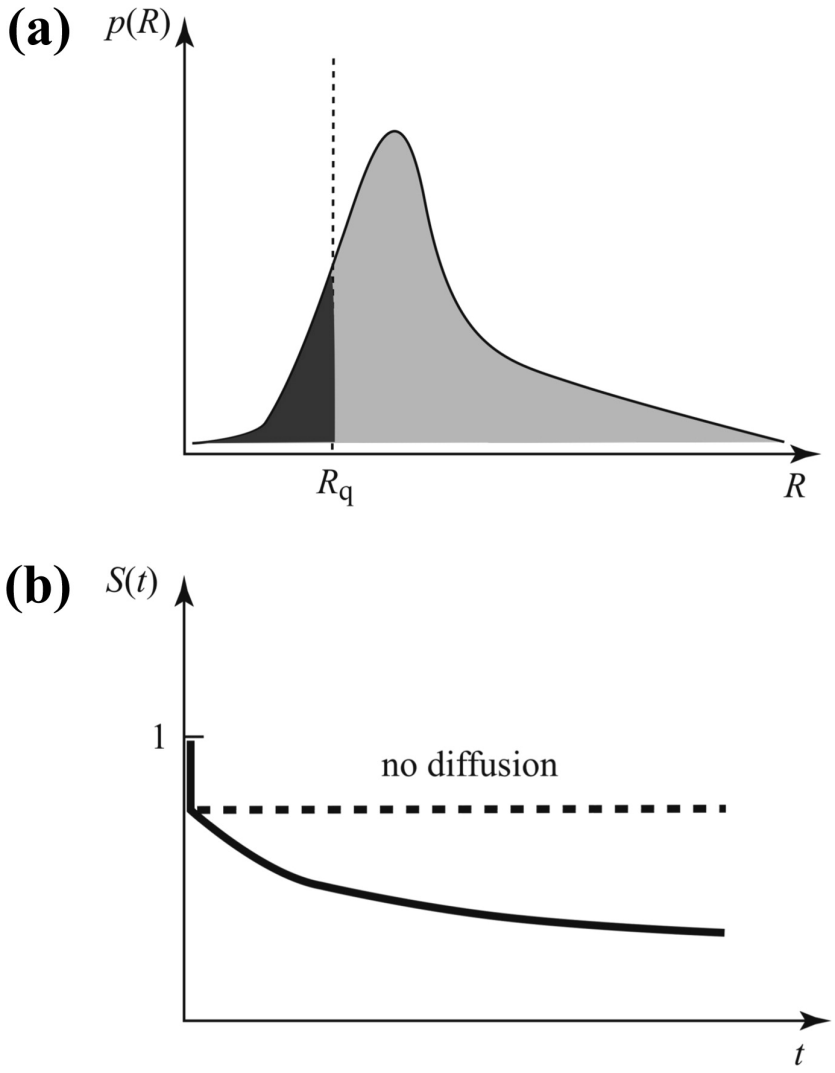


Figure 2.2: (a) In the contact quenching model (CQM), fluorescence quenching takes place instantaneously for all fluorophores for which $R < R_q$. This instantaneously depletes the dark shaded area in the probability distribution of the end-to-end distance for the excited molecules. (b) In the CQM, the survival probability for the excited probe undergoes a sharp drop at short times due to fluorophores that are within R_q of the quencher during excitation. This is followed by diffusion-controlled decay as other fluorophores diffuse into the quenching radius.

The situation is different when the quenching rate is finite (no matter how small) at any finite R. In this case, the survival probability will decay to zero even in the absence of any polymer motion. Moreover, $S(t)$ will decay instantaneously to zero in the limit $k_0 \to \infty$. To see this, we write the exact expression for the survival probability [17, 38–40]:

$$S(t) = \left\langle \exp\left(\int_0^t k[R(t')]\, \mathrm{d}t' \right) \right\rangle \tag{2.6}$$

where the angular brackets denote averaging over the trajectories $R(t)$ of the polymer chain starting with the equilibrium distribution. Since the exponent is proportional to k_0 (cf. Eq. 2.1), the survival probability will vanish in the $k_0 \to \infty$ limit for any finite time t.

Given these arguments, we should expect that, depending on the extent to which the quenching rate is poorly approximated as a step-function of distance, we will not observe well-defined, diffusion-controlled kinetics in experimental measurements of polymer dynamics. In contrast to this argument, the two most commonly employed approximations for describing polymer loop closure kinetics, SSS and WF theory, predict a well-defined, diffusion-controlled rate for any quenching mechanism. WF theory, in particular, predicts k_{eff} to approach the diffusion-limited constant value k_D given by the expression [31]

$$k_D^{-1} = k_R^{-2} \int_0^\infty \langle \Delta k[R(t)] \Delta k[R(0)] \rangle \, \mathrm{d}t \tag{2.7}$$

where

$$k_R = \int_0^\infty k(R) p(R) \, \mathrm{d}R \tag{2.8}$$

is the reaction-controlled rate and

$$\Delta k(R) = k(R) - k_R \tag{2.9}$$

The factor k_0 cancels out in Eq. 2.7, thus giving a value that is independent of the intrinsic quenching rate. We note that the possibility of a breakdown of the WF approximation at large values of k_0 has been anticipated (although not further explored) by Wilemski and Fixman in their original work [31].

2.3.2 In the limits of slow and rapid diffusion, the effective time of quenching scales as k_0^{-1}

When the quenching rate $k(R)$ is a continuous function of R, quenching takes place even in the absence of polymer motion. For example, in the limit of very high solvent viscosity (when the polymer is essentially frozen), the survival probability is given by [38]

$$S(t) \approx \langle \exp(-kt) \rangle = \int_0^\infty \exp[-k(R)t] p(R)\, \mathrm{d}R \tag{2.10}$$

It is known [17, 27] that the kinetics are strongly non-exponential in this limit and so the survival probability $S(t)$ does not predict a single unique time scale for quenching. Nevertheless, it is possible to define an effective time scale τ_F for quenching in this limit by considering the first passage time

$$\tau_F \equiv k_F^{-1} = \int_0^\infty S(t)\, \mathrm{d}t = \int_0^\infty p(R) k^{-1}(R)\, \mathrm{d}R = \langle 1/k \rangle \tag{2.11}$$

which can be estimated analytically if we assume that the polymer obeys Gaussian statistics such that $p(R)$ is given by

$$p(R) = 4\pi R^2 \left(\frac{3}{2\pi l^2 n} \right)^{3/2} \exp\left(-\frac{3R^2}{2l^2 n} \right) \tag{2.12}$$

where l is the Kuhn length and n is the number of statistical segments. Substituting Eq. 2.1 and 2.12 into Eq. 2.11, we find

$$k_F^{-1} = \frac{k_0^{-1}}{\sqrt{54\pi}}\{6g + \sqrt{6\pi}\exp(g^2/6)(3+g^2)[1+\mathrm{erf}(g/\sqrt{6})]\} \tag{2.13}$$

where

$$g = l\sqrt{n}/R_q = \sqrt{\langle R^2\rangle}/R_q \tag{2.14}$$

The dimensionless parameter g is a measure of the characteristic polymer length scale relative to the quenching length scale. For typical experimental peptide/DNA loop closure measurements [2, 22, 35], the quenching rate drops off very rapidly over the relevant polymer length scales. Therefore, the case $g \gg 1$ is of interest, where the following asymptotic expression is obtained:

$$k_F \approx \frac{3}{2}k_0 g^{-2}\exp(-g^2/6) \tag{2.15}$$

Note that k_F is proportional to k_0. Also note that the chain length and the quencher properties enter into the result only as the dimensionless combination g.

The parameter k_F provides an estimate of the rate of the "direct" process in which quenching occurs without polymer reconfiguration. As an example, consider the experiments of Lapidus et al. [22, 35], who measured the end-to-end collision rates of polypeptides containing $N = 5 - 20$ residues. The root-mean-square end-to-end distance for these polypeptides was estimated [35] to be in the range $\langle R^2\rangle^{1/2} \approx 1.11 - 2.22$nm. Using the value [41] $R_q = 0.025$nm, g corresponds to $44 - 100$, and according to Eq. 2.13,

$k_F = (5 \times 10^{-573} - 5 \times 10^{-146})k_0$. This leads, under the "no diffusion" assumption we are exploring here, to astronomically long quenching time scales for any reasonable k_0.

In the opposite limit, the limit of rapid diffusion, the overall rate is given by Eq. 2.8, which, again, is proportional to k_0. For Gaussian chains described by Eq. 2.12,

$$k_R = \frac{k_0}{\sqrt{54\pi}}\{-6g + \exp(g^2/6)(3+g^2)\sqrt{6\pi}(1 - \mathrm{erf}[g/\sqrt{6}])\} \tag{2.16}$$

Here, too, the chain length and quenching length dependencies enter into this result only through the dimensionless parameter g. Finally, we note that the rate k_R provides an upper bound on the observed collision rate k_{eff} while k_F gives a lower bound [17, 38].

2.3.3 Comparison of simulated rates with the Wilemski-Fixman approximation

Having discussed the limiting conditions of slow and rapid diffusion, we now turn to the general case. Here, however, we do not report an analytical solution. Instead, we report our results for the overall rate k_{eff} as a function of the intrinsic quenching rate k_0 obtained using Langevin dynamics simulations of a simple bead-and-spring polymer model (see Section 2.6 for technical details of the simulations). For the EQM, we find that at very low values of k_0 the reaction-controlled limit takes place and the dependence is linear (Fig. 2.3a), in accord with the WF approximation given by Eqs. 2.3, 2.7, and 2.8. At higher, though still low, values of k_0, the WF approximation remains accurate. As

noted above, we call this regime I. As the intrinsic quenching rate is further increased, however, the WF approximation predicts k_{eff} to approach a constant plateau value k_D (given by Eq. 2.7). Instead, the rate observed in our simulations continues to increase, although it exhibits a weaker dependence on k_0. We call this regime II. The departure of k_{eff} from the predictions of WF theory at high values of k_0 is observed regardless of the quenching length scale. For example, shortening R_q by $\sim$ 2-fold leads to a similar dependence (compare Fig. 2.3b with Fig. 2.3a).

Importantly, the departure from the WF theory cannot be explained simply by the onset of the slow diffusion limit described in Section 2.3.2. Indeed, in such a limit, the direct quenching mechanism not involving any polymer reconfiguration would dominate. However, the direct rate k_F estimated from Eq. 2.11 remains many orders of magnitude (e.g., a factor of 10^{16} for the data of Fig. 2.3a) smaller than the observed rate k_{eff} (Fig. 2.3c). Thus, regime II must result from a more subtle interplay between polymer dynamics and quenching kinetics.

As k_0 is further increased, the direct quenching mechanism will eventually become important. Indeed, the direct rate k_F is proportional to k_0, while the regime II rate exhibits a weaker, power-law dependence (see Section 2.3.4). Therefore, the curves representing these two dependences (Fig. 2.3c) must cross at some value of k_0. It is evident from Fig. 2.3c that the direct rate will become dominant only at values of k_0 that are many orders of magnitude higher than those explored in Fig. 2.3a. At such high values of k_0, the polymer may be

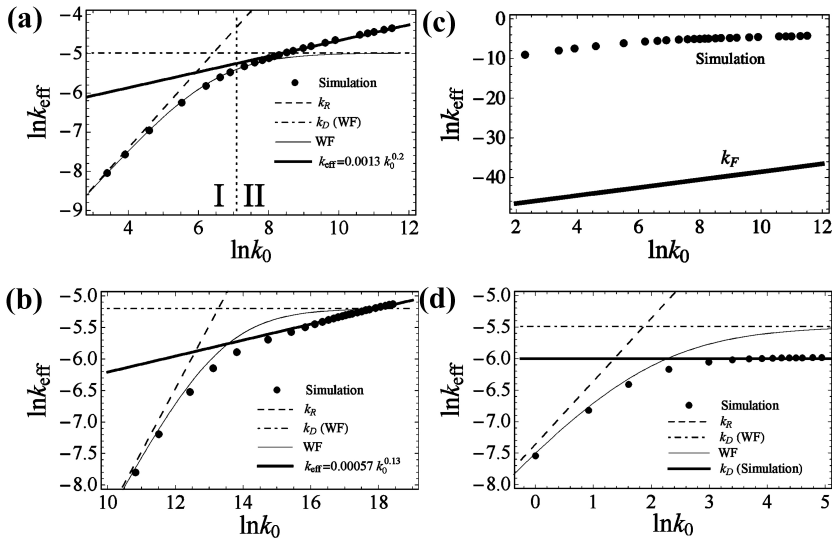


Figure 2.3: For the exponential quenching model (EQM), the Wilemski-Fixman (WF) theory (Eq. 2.3) is qualitatively incorrect in its prediction that the overall quenching rate k_{eff} attains a diffusion-controlled limit k_D (Eq. 2.7) that is independent of the intrinsic quenching rate k_0 at large values of k_0. In contrast, simulation results exhibit a power-law k_0 dependence. At low values of k_0, a reaction-controlled limit k_R is attained (Eq. 2.8), where k_{eff} is directly proportional to k_0. (a) Various theoretical predictions are compared with simulation data for $N = 14$ and $R_q = 0.2\sigma$, where σ is the polymer bond length. The dotted vertical line provides a rough boundary between regime I (where the WF approximation holds) and regime II (where the kinetics remains exponential and a power law holds). Regime III (where the polymer is effectively frozen) is well outside the range of k_0 plotted and occurs at $\ln k_0 \geq 20$. (b) Various theoretical predictions are compared with the simulation data for $N = 14$ and $R_q = 0.091\sigma$. (c) The direct rate k_F is many orders of magnitude lower than the actual quenching rate k_{eff} determined from simulations (same data as in Fig. 2.3a). Therefore, the onset of the direct quenching mechanism per se cannot explain the deviations from the WF behavior observed in regime II. (d) The WF theory predicts a qualitatively correct behavior of the overall quenching rate k_{eff} as a function of the intrinsic quenching rate k_0 for CQM, unlike the EQM case. Here, the WF approximation is compared with simulation data for $N = 14$ and $R_q = 2^{1/6}\sigma$. The value of R_q chosen here corresponds to the spatial range of the repulsive Lennard-Jones potential acting between monomers. The diffusion limited value k_D predicted by the WF theory agrees with the $k_0 \to \infty$ limit found from simulations to within a factor of ~ 1.7.

viewed as effectively frozen. In this limit, which we call regime III, the overall rate k_{eff} is not well defined because the kinetics (which are described by Eq. 2.10) are no longer exponential. Nevertheless, a corresponding time scale can still be defined as described in section Section 2.3.2. In the intermediate regime, regime II, the survival probability still shows exponential behavior and the rate constant k_{eff} can be estimated from the computed survival probability $S(t)$.

While the WF approximation fails to predict the dependence of k_{eff} on k_0 at high values of k_0 in the more realistic EQM case (shown in Fig. 2.3a and b), it remains qualitatively correct for any k_0 in the CQM case. Specifically, it predicts a diffusion-limited plateau $k_{eff} \to k_D$ at large values of k_0 (Fig. 2.3d). Although the numerical value of k_D predicted by the WF approximation differs from the exact value found in the simulations, they agree to within an order of magnitude, as previously discussed by Pastor et al [23]. Again, we emphasize that the difference between the performance of the WF approximation in the CQM and EQM cases (Fig. 2.3d versus Fig. 2.3a,b) is not just a matter of numbers: while in the former case it gives a qualitatively correct behavior for k_{eff} as a function of k_0, in the latter case, the WF approximation fails to capture even the qualitative observation that the quenching rate k_{eff} continues to rise as k_0 is increased.

2.3.4 Regime II: The power law dependence of k_{eff} on k_0 and viscosity

We now focus on the EQM case in regime II (see Fig. 2.3a), where exponential kinetics are still observed despite the fact that the rate significantly deviates from the WF approximation. In this regime (see Fig. 2.3a,b), the dependence of k_{eff} on k_0 is well described by a power law:

$$k_{eff} = Ak_0^{\delta} \tag{2.17}$$

We now show that the validity of Eq. 2.17 necessitates that the viscosity dependence of k_{eff} should also be a power law and not, as commonly assumed [22,28], inversely proportional. This can be seen from the following argument: Assuming an overdamped limit, the equations governing the polymer dynamics are of the general form

$$\gamma \frac{d\mathbf{r}}{dt} = \mathbf{F}(\mathbf{r}, t) \tag{2.18}$$

where γ is a friction coefficient (proportional to the solvent viscosity), $\mathbf{r}$ is the vector describing the configuration of the polymer, and $\mathbf{F}$ is the force (which includes the usual stochastic component). For a given polymer trajectory $\mathbf{r}(t)$, the survival probability $S(t) = \exp(-\int_0^t k[\mathbf{r}(t')]\,\mathrm{d}t')$ satisfies the equation

$$\frac{dS}{dt} = -k[\mathbf{r}(t)]S = -k_0\kappa[\mathbf{r}(t)]S \tag{2.19}$$

where $\kappa[\mathbf{r}] = \exp(-R/R_q)$ is a dimensionless function that describes the distance dependence of the quenching rate constant (see Eq. 2.1). Introducing dimensionless time units $\tau = k_0 t$, Eqs. 2.18 and 2.19 become

$$\gamma k_0 \frac{d\mathbf{r}(\tau)}{d\tau} = \mathbf{F}(\mathbf{r}, \tau) \tag{2.20a}$$

$$\frac{dS}{d\tau} = -\kappa[\mathbf{r}(\tau)]S(\tau) \tag{2.20b}$$

According to Eq. 2.20, the dynamics of the system depend on the "reduced" friction parameter γk_0 rather than on γ or k_0 individually. More importantly, the rate at which the survival probability decays to zero, measured in dimensionless units, also depends only on this reduced friction coefficient. Reverting to the original time units, we conclude that the overall rate must be of the form

$$k_{eff} = k_0 f(k_0\gamma) \tag{2.21}$$

where $f(x)$ is some yet unknown function. Thus, the γ dependence and the k_0 dependence of the rate are interrelated. We emphasize that the only assumption made in Eq. 2.21 is that the dynamics are overdamped.

We now can take advantage of the fact that in the regime of interest the k_0 dependence is a power law (Eq. 2.17). The function $f(k_0\gamma)$ must thus also be a power law. Specifically, to satisfy Eq. 2.17, it has to be given by $f(k_0\gamma) \propto (k_0\gamma)^{\delta-1}$, leading to

$$k_{eff} = k_0^{\delta}\gamma^{-(1-\delta)} \tag{2.22}$$

which is equivalent to Eq. 2.4. If, for example, $\delta = 0$, then $k_{eff} \propto 1/\gamma$ and the overall rate is inversely proportional to the friction coefficient (and viscosity) and is independent of k_0. This is the standard diffusion-controlled limit. If, on the other hand, $\delta = 1$, the rate is directly proportional to k_0 and is independent of viscosity. This dependence is observed both in the reaction-controlled and frozen limits (regimes I and III). The cases $\delta = 0$

and $\delta = 1$ represent two extremes, where the measured quenching kinetics are controlled, respectively, by the diffusive dynamics of the polymer and by the intrinsic quenching time scale. The fractional power law, $0 < \delta < 1$, represents the more general case where the two time scales compete [42]. Such fractional power-law viscosity dependence has been previously observed in the simulations of Srinivas and co-workers [27] for the case where $k(R)$ described fluorescence resonance energy transfer between the ends of a Rouse chain. Given the long length scale of FRET, this is perhaps not surprising. As we note below, however, fractional power-law dependencies have also been observed experimentally even when much more strongly distance-dependent quenching mechanisms have been employed.

What can be said about the proportionality coefficient in Eq. 2.22? There are two characteristic time scales in the problem. The first one is associated with the intrinsic quenching rate k_0^{-1}. The second one is the polymer's inherent time scale. Although polymer fluctuations generally exhibit a spectrum of time scales, the slowest one usually dominates long time dynamics. The relevant characteristic time can be estimated as [43, 44]

$$\tau_r = \frac{\gamma N \langle R^2 \rangle}{k_B T} \tag{2.23}$$

Note that we have omitted any numerical factors in our definition of τ_r. For example, the longest relaxation time within the Rouse model of polymer dynamics [43, 44] differs from the above equation by a factor of $3\pi^2$. Eq. 2.22 suggests that the exponent δ describes the partitioning of the overall relaxation

time between these two time scales. In other words, we can rewrite Eq. 2.22 in the form

$$k_{eff} = a k_0^{\delta} (1/\tau_r)^{1-\delta} \tag{2.24}$$

where a is a dimensionless coefficient. If R_q is longer than the polymer's Kuhn length, a should depend only on the ratio of the polymer's characteristic length scale $\langle R^2 \rangle^{1/2}$ and R_q. In other words, a is a function of the parameter $g = \langle R^2 \rangle^{1/2} / R_q$ defined in Section 2.3.2. Remarkably, most known analytic approximations to k_{eff} are limiting cases of Eq. 2.24. For example, it is straightforward to check that the SSS approximation [13, 22] for CQM is obtained by assuming $\delta = 0$ and $a = c/g$, where c is a numerical constant. In contrast, Doi [20] and Friedman and OShaughnessy [21] predicted that k_{eff} should be proportional to $1/\tau_r$. Their result is recovered if one assumes that $\delta = 0$ and a is a constant. Likewise, the expressions for the quenching rate in the limits of rapid and slow polymer diffusion (Eqs. 2.13 and 2.16) are also of the form of Eq. 2.24, where $\delta = 1$. These, however, are all limiting cases that do not rigorously describe real experimental systems. In the following sections, we will explore how the exponent δ and the proportionality coefficient a depend on the properties of the experimental system at hand.

2.3.5 The dependence of the exponent δ on the properties of the polymer and probe

While the viscosity dependence of the rate k_{eff} is readily measured via experiment, its dependence on k_0 is not, since the value of k_0 is generally

fixed by the physical properties of the probes employed. From a computational perspective, however, it is actually quite easy to determine δ from the k_0 dependence of k_{eff}, more so, in fact, than from its friction coefficient dependence. This is because a single polymer trajectory obtained at one value of the friction coefficient is sufficient to compute k_{eff} for any value of k_0. In view of Eq. 2.22, both approaches contain the same information and yield the same value of δ. Indeed, we have verified the validity of Eq. 2.22 by performing simulations over a range of viscosities (data not shown).

The value of the exponent δ generally depends on the specific polymer and on the probes employed to monitor its kinetics. Our goal here is to explore this dependence and to establish simple estimates of δ that would be applicable for various experimental conditions. As R_q decreases and quenching becomes increasingly short ranged, the exponent δ also decreases (Fig. 2.4a). This is consistent with previous simulation results [27] obtained for the kinetics of fluorescence resonance energy transfer between the ends of an ideal Rouse chain. In the case of infinitely short-range quenching ($R_q \to 0$), the $1/\eta$ scaling of the diffusion-controlled limit is recovered, which is intuitively expected, as in this limit quenching occurs upon contact ($R \approx 0$) and the difference between the CQM and EQM disappears. In practice, of course, this limit is of only academic interest, since the probability of such a contact would become vanishingly small.

In order to establish the dependence of δ on the properties of the polymer itself, we have also performed simulations for chains of different length N.

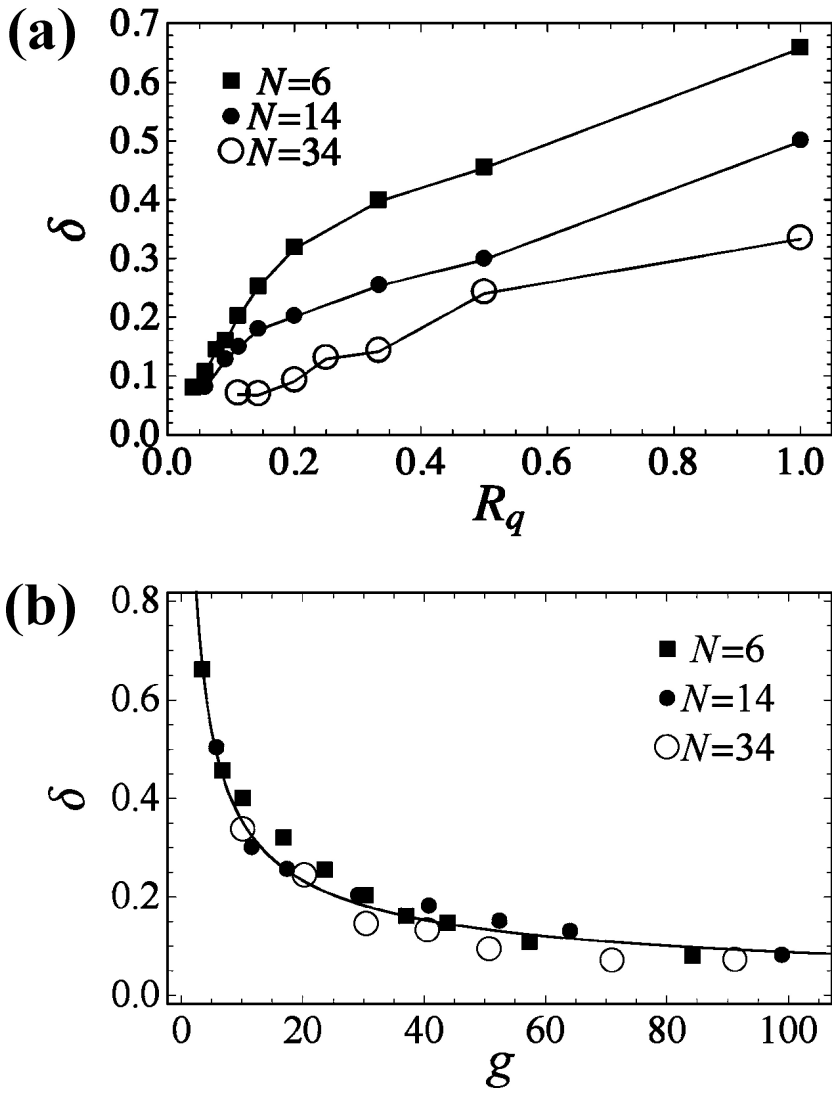


Figure 2.4: (a) The scaling exponent δ varies strongly as a function of the length scale of quenching and the chain length. (b) The scaling exponent δ approaches a universal curve when plotted as a function of the dimensionless parameter $g = \langle R^2 \rangle^{1/2} / R_q$. The solid line represents a power-law fit of this curve given by the equation $\delta \approx 1.416 g^{-0.602}$

These simulations show that the exponent δ decreases with increasing chain length (Fig. 2.4a).

The above findings regarding the dependence of δ on N and R_q were obtained for an idealized bead-and-spring polymer system. We can employ them, however, to estimate the numerical values of the exponent δ for various experiments without having to use a more realistic polymer model in each case. In the limit of sufficiently long chains and for the values of R_q that are larger than the Kuhn length of the chain, this is to be expected; in this limit, only two length scales are relevant, the quenching length scale R_q and the polymer's characteristic dimension $\langle R^2 \rangle^{1/2}$. Thus, the dimensionless exponent δ should depend on the ratio of the two, $g = \langle R^2 \rangle^{1/2} / R_q$, as it does in the two limiting cases discussed in Section 2.3.2. Fortuitously, a near universal relationship between δ and g holds even in the (more experimentally relevant) regime where R_q is comparable to or smaller than the chain Kuhn length. Indeed, when the data of Fig. 2.4a is replotted as δ versus g, it falls close to a single curve (Fig. 2.4b). The rather small deviations from a universal relationship presumably arise from finite size effects. The (nearly) universal curve $\delta(g)$ computed here can be used to analyze a variety of experimental systems as long as enough is known about them to estimate the value of g.

2.3.6 The dependence of the proportionality coefficient A on the properties of the polymer and probe

The proportionality coefficient A between k_{eff} and k_0^δ (cf. Eq. 2.17) strongly depends on chain length and R_q and spans several orders of magnitude, as these parameters are varied in our simulations (Fig. 2.5). As suggested by Eq. 2.24, we can write this proportionality coefficient as $A(R_q, N) = a(1/\tau_r)^{1-\delta}$, where τ_r is the polymer's inherent characteristic time (Eq. 2.23). The dimensionless parameters a and δ should become universal functions of the dimensionless parameter g defined in Eq. 2.14 in the asymptotic limit where both R_q and $\langle R^2 \rangle^{1/2}$ are much larger than the Kuhn length. Consistent with this, for the largest values of R_q studied here, the dependence $A(R_q, N)$ is well described by Eq. 2.24 with a constant value of $a \approx 0.5$ (Fig. 2.5). As R_q is decreased, deviations from the simple scaling formula of Eq. 2.24 are observed, although this formula still correctly describes the qualitative trend for A to decrease with decreasing R_q or with increasing chain length. These deviations are not surprising, since in this range of parameters we do not expect the parameter a to be constant or, more generally, to be any universal function of g. Nevertheless, Eq. 2.24, with $a \approx 0.5$, predicts the overall quenching time scale to within less than an order of magnitude across the entire range of R_q and N studied here.

In the limit $R_q \to 0$, our results suggest that A should approach a finite value (Fig. 2.5). This may appear counterintuitive, since one naturally expects the overall quenching rate to vanish in this limit. It is however important to

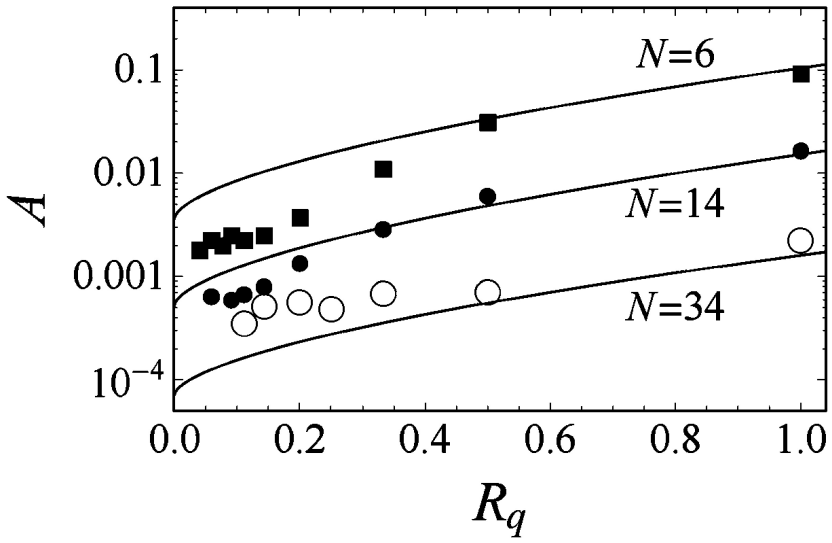


Figure 2.5: The proportionality coefficient A in the relationship $k_{eff} = Ak_0^\delta$ varies strongly as a function of the length scale of quenching and polymer length. Much of this variation is captured by the approximate formula $A = a(1/\tau_r)^{1-\delta}$, where τ_r is the polymer's inherent relaxation time (Eq. 2.23), $a = 0.5$, and the scaling exponent δ is estimated using the universal curve $\delta(g)$ of Fig. 2.4b. The predictions of this formula are plotted as solid lines.

remember that A represents the proportionality coefficient between k_{eff} and k_0^δ in the range of k_0 values such that the power law holds (i.e., in regime II). If R_q is decreased, the range of k_0 where the power law is valid gets shifted toward larger values (cf. Fig. 2.3a and b). On the other hand, k_{eff} indeed vanishes in the limit $R_q \to 0$ for any fixed value of k_0.

2.4 Comparison with experimental data

Fractional viscosity dependence has been observed by Bieri et al. [2] in their measurements of the end-to-end collision rates in short polypeptides.

They found that $k_{eff} \propto \eta^{\delta-1}$ with δ close to 0.2 except for the shortest peptide studied, for which δ is ~ 0.04. Since the chains used in that study were rather short ($3 \leq N \leq 10$), we do not necessarily expect the universal behavior described here (Fig. 2.5b) to hold. Nevertheless, if we assume it to be true, the value $\delta = 0.2$ corresponds to $g = 20 - 40$. Estimating $\langle R^2 \rangle^{1/2} \sim 1.5$ nm for a polypeptide with $N = 10$ bonds, this corresponds to $R_q \sim 0.4 - 0.8$ Å.

In contrast to the nonlinear relationship predicted here, and reported by Bieri et al. [2], Lapidus et al. [22] report a linear relationship between k_{eff}^{-1} and viscosity and interpret the intercept as the value associated with the reaction-controlled rate k_R (see Eq. 2.3). In Section 2.3.2, however, we have estimated $g = 44 - 100$ for their system, thus corresponding to $\delta < 0.15$. Although a power-law fit was not attempted by Lapidus et al., it appears that it would be difficult to distinguish between the $k_{eff}^{-1} \propto \eta^{1-\delta} = \eta^{0.85} - \eta^{0.9}$ predicted here and the linear k_{eff}^{-1} vs η dependence expected on the basis of the WF or SSS approximations.

As yet another example, we have recently studied the end-to-end collision rates of single-stranded DNA [37]. For a 15-base polythymine chain, deviations from a linear dependence of k_{eff}^{-1} on the viscosity are readily apparent (Fig. 2.6). Specifically, a power-law fit $k_{eff}^{-1} \propto \eta^{1-\delta}$ employs the same number of adjustable parameters as the linear fit predicted by the CQM and provides a statistically significantly improved fit (Fig. 2.6). The value of δ obtained from this fit, $\delta \approx 0.524$, is higher than the value estimated using the universal curve (Fig. 2.4b). Specifically, assuming a Kuhn length of 4 nm for

a single-stranded DNA [45] and using a bond length of $\sigma = 0.34$ nm for single-stranded DNA [46], we estimate $\langle R^2 \rangle^{1/2} \approx (N\sigma l)^{1/2} \approx 4.6$ nm. Using the value [47] $R_q = 0.725$ Å, we estimate $g \sim 60$, which corresponds to $\delta \approx 0.14$ (Fig. 2.4b). The discrepancy between the observed and predicted values of the exponent, however, is not surprising because the contour length of the chain is not much longer than the Kuhn length and so the experimental conditions are rather far from the flexible chain assumed in this analysis. Moreover, our recent study [37] demonstrates that electrostatic interactions within rather short single-stranded DNAs such as this lead to significant finite size effects that cannot be captured by simple analytic polymer models.

Given the approximations still inherent in our model, and its inability to model the specific details of each of these experimental setups, even the limited degree of correlation observed between theory and experiment appears heartening. Moving forward, a more realistic model for the end-to-end collision dynamics would also have to take into account the effect of flexible linkers connecting DNA to the probes [37]. One could attempt to capture the effect of the linkers with a naive model, in which $k(R)$ represents the overall rate of two processes: The first one consists of a diffusional approach between the fluorophore and the quencher groups subject to the constraint imposed by tethering those probes to the ends of the DNA chain. The second process is the fluorophore quenching. The effect of the flexible linkers is thus to effectively increase the spatial range of the quenching process. The characteristic length scale for $k(R)$ would then be comparable with the typical end-to-end distances

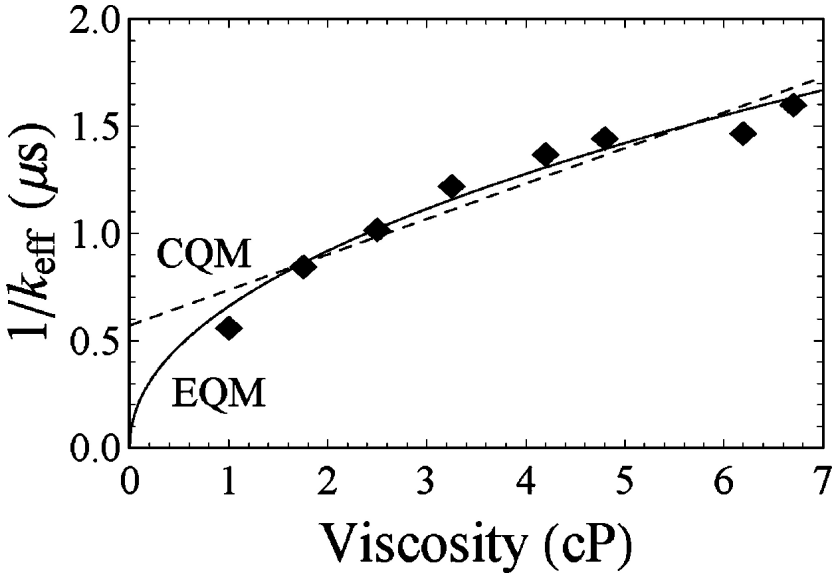


Figure 2.6: The EQM predicts a nonlinear, power-law relationship between solvent viscosity and the end-to-end collision rate of a polymer (here a 15-base single-stranded DNA comprised of polythymine) when probed using a long-lived lumiphore (ruthenium tris(bipy)) quenched via electron transfer (to methyl-ethyl-viologen). While the difference in the two fits may appear subtle, both models entail only two fitted coefficients and thus the differences in the residuals of the two fits are statistically significant ($R^2 = 0.96$ versus 0.90 for the EQM and CQM, respectively). The experimental measurements were performed by our collaborators T. Uzawa and K. W. Plaxco [12].

for the linkers and would be longer than the assumed value of R_q. It then appears plausible that the presence of the linkers would lead to an effectively lower value of g and thus a higher value of δ, consistent with our experimental findings.

2.5 Concluding Remarks

Experimental methods for measuring end-to-end collision dynamics of polypeptides and DNA commonly utilize the quenching of an optically excited probe attached to one chain end upon its collision with a quencher attached to the other end. The main message of this study is that the time scales of quenching and of the polymer conformational dynamics are inherently inseparable in such measurements. As a consequence, even their qualitative theoretical interpretation requires explicit consideration of the quenching mechanism of the experimentally employed probes.

Mathematically, the inseparability of the time scales of quenching and of the polymer conformational dynamics arises from the distance dependence of the quenching rate $k(R)$, which is a decreasing function of the distance between the chain ends. Even if this rate is so rapid that quenching is effectively instantaneous at sufficiently short distances, the time scale of quenching becomes comparable with that of polymer reconfiguration at larger values of R. Using simple scaling arguments, we have then shown that if the polymer dynamics are overdamped then the overall quenching time scale τ is given by

$$\tau^{-1} = k_{eff} = k_0 f(k_0 \tau_r) \tag{2.25}$$

where k_0 is the intrinsic quenching rate, τ_r is the polymer's intrinsic reconfiguration time (Eq. 2.23), and $f(x)$ is some function that depends parametrically on the chain length and R_q. Eq. 2.25 is exact and may serve as a starting point for developing various useful approximations for k_{eff} through mapping

out of the function $f(x)$.

Simulations of model polymer chains performed here indicate that in the regime of moderately high values of k_0 where the quenching kinetics remains exponential this function is close to a power law, $f(x) \propto x^{\delta-1}$, thus necessitating that the dependence of k_{eff} on both viscosity and intrinsic quenching rate be a power law. The fractional power-law viscosity dependence is consistent with earlier observations for other systems [27,39,40,42,48]. According to those studies, such a dependence is a common signature of the competition between the time scales of diffusion and reaction kinetics. Although our own simulations here employed a particular functional form for the quenching rate $k(R)$, in view of those studies, we anticipate that the power law found here for EQM would be more general and would not require an exponential dependence of Eq. 2.1. In fact, Srinivas et al. [27] report a fractional power law in the viscosity dependence for $k(R) = k_0/[1+(R/R_0)^6]$, where R_0 is a constant.

Using dimensional analysis, we have further shown that the dependence of the measured loop closure rate on the properties of the probes and on the polymer length can be recast as the scaling law given by Eq. 2.24, where $a = a(g)$ and $\delta = \delta(g)$ are universal in the asymptotic limit where the chain dimensions and the quenching length are much longer than the monomer size. We also found that near universality persists even outside this asymptotic limit, where the quenching length is comparable with the monomer size.

The scaling exponent δ describes how the overall time scale of quenching is partitioned between the inherent time scales of quenching and polymer

dynamics. If, for example, $\delta = 0$, then polymer dynamics is rate limiting and k_{eff} becomes independent of the quenching kinetics and inversely proportional to viscosity. In the opposite extreme, $\delta = 1$ and k_{eff} is independent of the polymer dynamics and is proportional to k_0, as in the reaction-controlled limit (where polymer dynamics are infinitely rapid) and the limit where the polymer is effectively frozen. Moreover, standard analytic approximations are recovered as various limiting cases of the scaling relationship of Eq. 2.24. For example, the SSS approximation [22] corresponds to $a \propto g^{-1}$, $\delta = 0$, while the Doi result [20] (described above) is obtained if $\delta = 0$ and a is constant.

The scaling prediction of Eq. 2.24 allows us to estimate the error incurred when using the standard approximations (such as the CQM) that do not explicitly take into account the quenching mechanism and predict a diffusion-controlled limit that is independent of k_0. This approximation corresponds to choosing $\delta = 0$ and underestimates the more accurate (EQM-predicted) rate by a factor of $k_{eff}^{EQM}/k_{eff}^{CQM} \sim (k_0\tau_r)^{\delta}$. When δ is close to zero and/or when the intrinsic quenching rate k_0 does not significantly exceed τ_r^{-1}, then the interpretation of the experimental data in terms of the approximate CQM does not lead to significant errors.

Finally, we note that the characteristic polymer reconfiguration time scale τ_r has attracted considerable attention from experimentalists in the context of the dynamics of unfolded proteins (see, e.g., refs [9–11]). If the intrinsic quenching rate k_0 can be measured independently, Eq. 2.24 provides another method for measuring τ_r.

2.6 Appendix

Langevin dynamics simulations of polymer chains. We used a simple polymer model that consisted of $N+1$ beads of mass m connected by N springs. The interaction potential between a pair of beads separated by a distance r consists of a harmonic bond interaction for two covalently linked beads

$$V_{bond}(r) = k_{bond}(r-\sigma)^2/2 \tag{2.26}$$

and a repulsive Lennard-Jones potential between nonbonded beads

$$V_{nonbonded}(r) = 4\epsilon[(\sigma/r)^{12} - (\sigma/r)^6]\theta(2^{1/6}\sigma - r) \tag{2.27}$$

Here, the parameter ϵ sets the energy scale, σ represents the equilibrium bond distance, $k_{bond} = 100\epsilon/\sigma^2$, and $\theta(x)$ is the Heaviside step function used to truncate the attractive part of the Lennard-Jones potential. The dynamics of each bead was described by the Langevin equation

$$m\frac{d^2\mathbf{r}_i}{dt^2} = -\frac{\partial V}{\partial \mathbf{r}_i} - \gamma\frac{d\mathbf{r}_i}{dt} + \mathbf{R}_i(t) \tag{2.28}$$

where $\mathbf{r}_i$ is the position of the bead, V is the total potential, γ is the friction coefficient, and $\mathbf{R}_i(t)$ is the random force satisfying the fluctuation-dissipation theorem. All simulations reported here were performed for $\gamma = 2.0(\sigma^2/m\epsilon)^{-1/2}$ and a temperature of $T = 1.0\epsilon/k_b$. In reporting our data, we use the bond length σ as the unit of distance and $\tau_0 = (m\sigma^2/\epsilon)^{1/2}$ as a unit of time so that the reported rate constants are normalized by the factor $(m\sigma^2/\epsilon)^{-1/2}$.

To obtain the overall quenching rate, we computed the survival probabilities $S(t)$ defined by Eq. 2.6 using $R(t)$ from Langevin trajectories. In the range of parameters reported here, $S(t)$ was found to be close to exponential, $S(t) \approx \exp(-k_{eff}t)$, thus yielding the values of the overall quenching rate k_{eff}.

Chapter 3

End-to-surface reaction dynamics of a single surface-attached DNA or polypeptide

3.1 Abstract

The dynamics of surface-attached polymers play a key role in the operation of a number of biological sensors, yet its current understanding is rather limited. Here we use computer simulations to study the dynamics of a reaction between the free end of a polymer chain and a surface, to which its other end has been attached. We consider two limiting cases, the diffusion-controlled limit, where the reaction is accomplished whenever the free chain end diffuses to within a specified distance from the surface, and the reaction-controlled limit, where slow, intrinsic reaction kinetics rather than diffusion of the chain is rate limiting. In the diffusion-controlled limit, we find that the overall rate scales as N^{-b}, where N is the number of monomers in the chain and $b \approx 2.2$ for excluded volume chains. This value of the scaling exponent b is close to that derived from a simple approximate theory treating the dynamics of the chain end relative to the surface as one-dimensional diffusion in an effective potential. In the reaction-controlled limit, the value of the scaling exponent b is close to 1. We compare our findings with those for the related (and better studied) problem of end-to-end reactions within an unconstrained polymer

chain and discuss their implications for electrochemical DNA sensors.

3.2 Introduction

Surface-attached biopolymers are employed in many emerging biotechnological applications, such as DNA microarrays [49], as well as in experimental biophysical methods. For example, many optical single-molecule techniques require the molecule of interest to be immobilized on a surface [50, 51]. Yet the current understanding of how surface confinement alters the dynamics and the structure of biomolecules is rather limited. In particular, theoretical work on this subject has been limited to only a handful of recent studies [52–59]. Shang and Geva, for example, have studied how tethering a freely jointed homopolymer [53] or a polypeptide [56] to either an attractive or repulsive surface affects their conformational ensemble and dynamics under varying solvent quality. Several studies of surface attachment effects on protein stability and folding kinetics have been reported [57,58,60,61]. The dynamics of a DNA tethered to a surface and subjected to a shear flow has also been studied [59]. Finally, molecular dynamics simulations of short surface-attached biopolymers have recently been reported [54, 55].

Here, we study the dynamics of a reaction occurring between one end of a polymer chain and a surface, to which the other end is attached (Fig. 3.1). Our interest in this problem is motivated in part by its particular application to the design of electrochemical DNA (E-DNA)sensors [8]. An E-DNA sensor is used to detect target DNA sequences and consists of an array of single-

stranded DNA (ssDNA) probes. One end of each molecule is tethered to an electrode surface, while the free end is labeled with a redox group. When no complementary target DNA is present, the redox-label can collide with the electrode surface, resulting in electron transfer that can be detected by measuring the resulting Faradic current [7]. When the probes hybridize with the target DNA, the resulting double-stranded DNA (dsDNA) segment is much stiffer than the unhybridized probe, so the collisions of its end with the surface become less likely. The ensuing change in the Faradic current thus can be linked to the concentration of the target DNA.

A quantitative description of E-DNA sensors requires a theory for the dependence of the electron transfer rate on polymer properties, such as persistence length and contour length. In general, this problem is rather complex because the chain dynamics are affected by electrostatic interactions of the DNA probe with the electrode, with neighboring DNAs and with itself. Moreover, a comprehensive theory should account for the possibility of through-DNA electron transfer, in addition to the collisional mechanism described above. Our objective here is more modest: We try to gain initial understanding of reaction dynamics between the chain end and the surface by considering the relatively simple case of a random-coil-like polymer attached to a hard wall-like surface.

Although, to our knowledge, this problem has not been studied in the literature (with the exception of ref [62] , which considered the case of a short, stiff dsDNA that has to bend for its end to reach the surface), the related problem of end-to-end collisions within a freely diffusing polymer chain (Fig. 3.1)

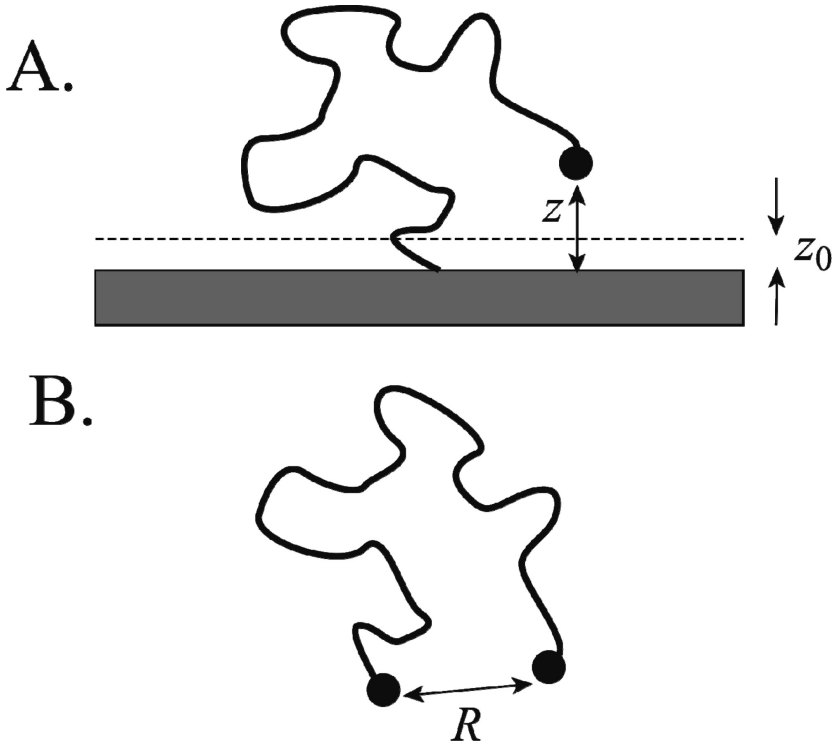


Figure 3.1: (a) Schematic picture of the problem studied here: One end of a polymer chain is attached to the surface. Reaction between the free chain end and the surface occurs inside a thin layer whose width is z_0. (b) A related problem of end-to-end collisions within a chain. Reaction occurs whenever the end-to-end distance is smaller than z_0.

has been studied rather extensively, both experimentally [2, 6, 22, 34–37, 63] and theoretically [3, 18, 20, 21, 23–26, 28–31, 64–72]. In particular, the scaling of the collision rate as a function of chain length has received considerable attention and still remains a somewhat unsettled issue [28]. While brute force computer simulations have recently provided a wealth of information about end-to-end collision dynamics [12, 23–25, 28], two approximate solutions, the Wilemski-Fixman approximation [31] and the Szabo-Schulten-Schulten (SSS)

theory [13], remain useful in interpreting experimental data. The SSS theory is particularly attractive, despite its known limitations [12, 23, 26], as it often allows for an analytically tractable solution. This theory assumes that the end-to-end dynamics of the polymer obeys a simple one-dimensional diffusion equation in the effective potential that correctly reproduces the distribution of end-to-end distances for the molecule.

In this paper, in addition to solving the problem of end-to-surface reaction dynamics via brute-force simulations, we also seek simple approximations to describe the results. We find that SSS theory is reasonably accurate as a description of the dependence of the rate, with which the free end of the polymer chain collides with the surface, on chain length and the contact distance z_0 defined such that the collision takes place only when the distance z between the free end and the surface is less than z_0 (Fig. 3.1). The SSS theory, as well as dimensional analysis, further suggest that, under the assumptions that the chain is long enough and that the contact distance is greater than the bond length but shorter than the root-mean-square (rms) distance from the chain end to the surface (i.e., $\sigma \ll z_0 < \langle z^2 \rangle^{1/2}$), the collision rate should be of a universal form:

$$k = t_R^{-1} f(z_0^2 / \langle z^2 \rangle)$$

Here t_R is the characteristic relaxation time of the polymer, and $f(x)$ is a universal function, which can be fairly accurately estimated from the SSS theory. We also study the opposite limit of reaction-controlled kinetics. In

this limit, slow, intrinsic reaction kinetics rather than diffusion of the chain is rate limiting. If such an intrinsic process takes place with a characteristic rate constant k_0, then we find the overall rate to obey another universal scaling relationship,

$$k = k_0 g(z_0^2 / \langle z^2 \rangle)$$

We note, however, that some of these findings may be model dependent. In particular, the model adopted here accounts for excluded-volume interactions within the polymer chain but, similarly to many other common models of end-to-end collision dynamics [22–25, 28], neglects hydrodynamic interactions. It is known that, in the asymptotic limit of very long chains, hydrodynamic interactions can qualitatively change the physics of the end-to-end chain reactions. For example, when both excluded volume and hydrodynamic interactions are present, the diffusion-controlled limit becomes unattainable, and reaction-controlled kinetics takes place even when the intrinsic reaction kinetics is very fast [21, 66, 67, 69, 70, 72].

Subsequent sections of this paper will be presented as follows: Section 3.3 will discuss the model and methods used in our study. Sections 3.4 and 3.5 describe our results for the diffusion-controlled rate and the reaction-controlled rate, respectively. In Section 3.6, we compare our results with those obtained for the kinetics of reactions taking place between ends of an unconstrained polymer chain. Finally, Section 3.7 concludes with a discussion of the implications for DNA sensors.

3.3 Model and Methods

3.3.1 Langevin dynamics simulations of a surface-attached polymer chain

Our polymer model consists of $N+1$ beads of mass m connected by N springs, where the first bead is immobilized on the surface at $z = 0$. Here the coordinate z measures the distance from the surface as shown in Fig. 3.1. The pairwise interaction potential between adjacent beads separated by a distance r is taken to be harmonic:

$$V_{bond}(r) = k_{bond}(r - \sigma)^2/2 \tag{3.1}$$

while nonbonded beads interact via a repulsive Lennard-Jones potential representing excluded volume:

$$V_{non-bonded}(r) = 4\epsilon[(\sigma/r)^{12} - (\sigma/r)^6]\theta(2^{1/6}\sigma - r) \tag{3.2}$$

Here, σ is the equilibrium bond distance, ϵ is a characteristic energy scale, $k_{bond} = 100.0\epsilon/\sigma^2$, and θ is the Heaviside step function that truncates the attractive portion of the Lennard-Jones potential. The interaction between the surface and each bead (except the one that is immobilized), as a function of its distance z between the bead and the surface, is similarly represented by a repulsive Lennard-Jones potential:

$$V_{surface}(z) = 4\epsilon[(\sigma/z)^{12} - (\sigma/z)^6]\theta(2^{1/6}\sigma - z) \tag{3.3}$$

The dynamics of each bead is governed by the Langevin equation

$$m\frac{d^2\mathbf{r}_i}{dt^2} = -\frac{\partial V}{\partial \mathbf{r}_i} - \xi\frac{d\mathbf{r}_i}{dt} + \mathbf{R}_i(t) \tag{3.4}$$

where $\mathbf{r}_i(t)$ is the position of the bead, V is the total interaction potential, ξ is a friction coefficient, and $\mathbf{R}_i(t)$ is a random force satisfying the fluctuation-dissipation theorem. Our data is reported in reduced form where the units of length, energy, and time are σ, ϵ, and $\tau_0 = (m\sigma^2/\epsilon)^{1/2}$, respectively. Simulations were performed with a temperature of $T = 1.0\epsilon/k_B$. The value $\xi = 2.0(\sigma^2/m\epsilon)^{-1/2}$ chosen for the friction coefficient was large enough to ensure that the dynamics were overdamped (i.e., all reaction times reported here are proportional to ξ).

3.3.2 The kinetics of end-to-surface reaction

We assume that there is a first-order reaction between the chain end and the surface. This reaction can represent, for example, electron transfer in a DNA sensor or surface-modulated quenching of an optical probe attached to the chain end. The rate for the reaction, $k = k(z)$, depends on the distance z (Fig. 3.1).

Although the actual functional form of $k(z)$ depends on the specifics of the above reaction, here we will consider the simplest possible model, where $k(z)$ is nonzero only when the distance z falls below a certain contact distance z_0, i.e.,

$$k(z) = k_0\theta(z_0 - z) \tag{3.5}$$

where θ is the Heaviside step function. We note that the chosen form for $k(z)$, which vanishes identically for $z > z_0$, is somewhat special, resulting in a well-defined limit $k_0 \to \infty$, where the overall measured rate becomes independent

of the intrinsic rate k_0 (i.e., the diffusion controlled limit). This is not true in general [12]. Nevertheless, we adopt the simplest possible model of Eq. 3.5 to avoid the complications encountered in the more general case [12]. Simulations can of course be performed for any $k(z)$, if desired.

To describe the overall kinetics of the polymer-mediated reaction, we compute the survival probability that the reaction initiated at time t_0 has not happened by time $t_0 + t$ [17, 38–40]:

$$S(t) = \left\langle \exp\left(- \int_{t_0}^{t_0+t} k[z(t')]\, \mathrm{d}t' \right) \theta[z(t_0) - z_0] \right\rangle / \left\langle \theta(z - z_0) \right\rangle \tag{3.6}$$

The averaging, denoted by the angular brackets in Eq. 3.6, is over the canonical ensemble of initial polymer configurations at $t = t_0$, with any initial configurations that have already reacted (i.e., those with $z < z_0$) excluded by the step function in Eq. 3.6. Assuming ergodic behavior, the averages of Eq. 3.6 can also be computed from a single long polymer trajectory by performing a time average over the starting point t_0. In practice, we have performed both averages by running long trajectories on multiple processors and carrying out a time average for each. The mean first passage time for the reaction to take place is then computed as

$$\tau = \int_0^\infty S(t)\, \mathrm{d}t \tag{3.7}$$

We are particularly interested in two limits. In the limit $k_0 \to \infty$ the result becomes independent of k_0. This is the diffusion-controlled limit, where the reaction occurs instantaneously upon the chain end, reaching the distance z_0 from the surface. The resulting diffusion-controlled rate $k_D = \tau^{-1}$ is a

measure of collisions between the chain end and the surface. In practice, to compute k_D from Eq. 3.7, we use sufficiently high values of k_0 such that the mean first passage time no longer depends on k_0. Note, however, that, although the diffusion-controlled limit is well-defined for the model studied here, it is not necessarily the case in general. For example, lack of diffusion-controlled limit has been predicted in loop formation within very long chains [21, 66, 67, 69, 70, 72], as well as the case of a more general spatial dependence of the reaction rate $k(z)$ [12].

In the opposite limit $k_0 \to 0$, the survival probability can be written as [17, 22, 38] $S(t) \approx \exp[-\langle k \rangle t] \equiv \exp(-k_R t)$, where

$$k_R = \langle k \rangle = \int_0^\infty k(z) p(z)\, \mathrm{d}z = k_0 \int_0^{z_0} p(z)\, \mathrm{d}z \tag{3.8}$$

is the reaction controlled rate, which can be obtained by averaging $k(z)$ over the equilibrium probability distribution of the distance between the chain end and the surface. While $\tau = k_R^{-1}$ can be computed in this limit again by considering the decay of an appropriately defined survival probability, as in Eq. 3.6 and Eq. 3.7, it is easier to calculate the reaction-controlled rate directly from Eq. 3.8. Finally, for intermediate values of k_0, one can interpolate between the diffusion- and reaction-controlled limits using the formula [22] $k = (k_D^{-1} + k_R^{-1})^{-1}$.

3.3.3 Estimating the diffusion-controlled rate for a Gaussian chain using the SSS theory

In Section 3.4, our numerical results will be compared with the analytically solvable SSS model [13]. In this model, the free end of the chain is viewed as diffusing in a potential of mean force of the form $G(z) = -k_B T \ln p(z)$, where $p(z)$ is the equilibrium distribution of the end-to-surface distance. The first passage time to reach the contact distance z_0 and the corresponding diffusion-controlled rate constant k_{D-SSS} are then given by [13]

$$\tau_{D-SSS} = k_{D-SSS}^{-1} = \int_{z_0}^{\infty} (D_{eff} p(z))^{-1} \left[\int_z^{\infty} p(s)\, \mathrm{d}s \right]^2 \mathrm{d}z \tag{3.9}$$

where D_{eff} is the effective diffusion coefficient along the coordinate z. Equation 9 can be evaluated analytically for polymers obeying Gaussian statistics. Indeed, the probability distribution $p(z)$ for the distance between the free end of a Gaussian chain and a surface, to which its other end is attached, is given by [73]

$$p(z) = \left(\frac{3}{N\sigma l_K} \right) z \exp\left(-\frac{3z^2}{2N\sigma l_K} \right) \tag{3.10}$$

where l_K is the Kuhn length [43] (thus the mean squared end-to-end distance measured in the absence of the surface is equal to $N\sigma l_K$). Inserting Eq. 3.10 into Eq. 3.9, we find

$$\tau_{D-SSS} = \frac{N\sigma l_K}{6D_{eff}} \Gamma\left(0, \frac{3z_0^2}{2N\sigma l_K} \right) = \frac{\langle z^2 \rangle}{4D_{eff}} \Gamma\left(0, \frac{z_0^2}{\langle z^2 \rangle} \right) \tag{3.11}$$

where Γ is the incomplete gamma function and $\langle z^2 \rangle = \int_0^{\infty} z^2 p(z)\, \mathrm{d}z = (2/3) N\sigma l_K$ is the mean square distance between the chain end and the surface.

In order to use Eq. 3.11 to predict the diffusion-controlled collision rate in our model, we need two parameters, the Kuhn length l_K and the effective diffusion coefficient D_{eff}. The former can be estimated by fitting the mean squared end-to-end distance of the chain (in the absence of the surface) with $N\sigma l_K$. The latter can be estimated by comparing the behavior of the autocorrelation function of the end-to-surface distance $z(t)$ for polymer model with that expected for the one-dimensional diffusion model [74]. Specifically, using a harmonic approximation for the potential of mean force

$$G(z) = -k_B T \ln p(z) = G(z_m) + (1/2)G''(z_m)(z - z_m)^2 \tag{3.12}$$

where $z_m = (N\sigma l_K/3)^{1/2}$ corresponds to the maximum of $p(z)$, one can estimate the effective diffusion coefficient as [74]

$$D_{eff} = \frac{k_B T}{G''(z_m)} \frac{C(0)}{\int_0^\infty C(t)\,\mathrm{d}t} \tag{3.13}$$

where

$$C(t) = \langle z(t)z(0)\rangle - \langle z\rangle^2 \tag{3.14}$$

The correlation function of Eq. 3.14 can be often approximated by a single exponential,

$$C(t) = C(0)\exp(-t/t_0) \tag{3.15}$$

We have directly tested Eq. 3.15 by simulations and found that it always holds well. If we further assume that our Gaussian polymer obeys Rouse dynamics (i.e., Brownian dynamics, with hydrodynamic effects neglected [44]), then we

expect that $C(t)$ would be dominated by the slowest Rouse relaxation mode so that $t_0 = t_R$, where

$$t_R = \frac{N^2 \sigma l_K \xi}{3\pi^2 k_B T} \tag{3.16}$$

is the Rouse relaxation time of the chain. Indeed, neglecting faster relaxation modes in the expression for $C(t)$ results in a fairly small error ($\approx 10\%$) when estimating the effective end-to-end diffusion coefficient of a free Rouse chain [74]. However, this result has to be modified to account for the fact that one end of the chain is fixed. This leads to a different boundary condition for the Rouse relaxation modes [75]. Consequently, for a Rouse chain with one end fixed, the corresponding slowest relaxation time is not t_R but $t_0 = 4t_R$ [75]. Using Eqs. 3.13, 3.14, 3.15, and 3.16, one finds

$$D_{eff} \approx \frac{k_B T}{G''(z_m) t_0} = \frac{N \sigma l_K}{6 t_0} = \frac{N \sigma l_K}{24 t_R} = \frac{\pi^2 k_B T}{8 N \xi} \tag{3.17}$$

Similarly to the free Rouse chain, D_{eff} of a constrained chain is inversely proportional to chain length [74].

3.4 Results: The diffusion-controlled rate k_D

3.4.1 Dependence of the diffusion-controlled rate on chain length N and the contact distance z_0

In the diffusion-controlled limit, the reaction kinetics are determined by the collisions of the free chain end with the surface, where a collision is defined as an event in which the chain end approaches the surface to within the contact distance z_0. The mean first passage time τ to such a collision is

related to the usual diffusion-controlled rate [13], $k_D = \tau^{-1}$. For sufficiently small values of z_0 and/or long chains, the chain length dependence of k_D is well approximated by a power law

$$k_D \propto N^{-b} \tag{3.18}$$

where the value of the exponent b is close to 2.16 (Fig. 3.2, the case $z_0 = 2^{7/6}\sigma$). Deviations from a power law dependence become significant as z_0 is increased and becomes comparable with the typical distance z between the surface and the chain end.

The SSS rate k_{D-SSS} (Eq. 3.11 and Eq. 3.17), while being off roughly by an order of magnitude for the longest chains studied, reproduces remarkably well both the scaling exponent b in the limit of small z_0 and the deviations from the power law observed at larger values of z_0 and/or for short chains.

Likewise, the dependence of the diffusion-controlled rate on the contact distance z_0 is well captured by the SSS theory (Fig. 3.3). Although this dependence appears to be well approximated by an exponential function $k_D \propto \exp(az)$ (solid lines in Fig. 3.3), this behavior may be accidental. Indeed, the SSS prediction of Eq. 3.11 also appears to behave exponentially in the same range of z_0, while the actual analytical formula is inversely proportional to an incomplete Gamma function involving z_0^2 and is not an exponential function of z_0.

When extrapolated to $z_0 \to 0$, our numerical data for k_D appears to predict a finite value of the diffusion-controlled rate. This simple extrapolation

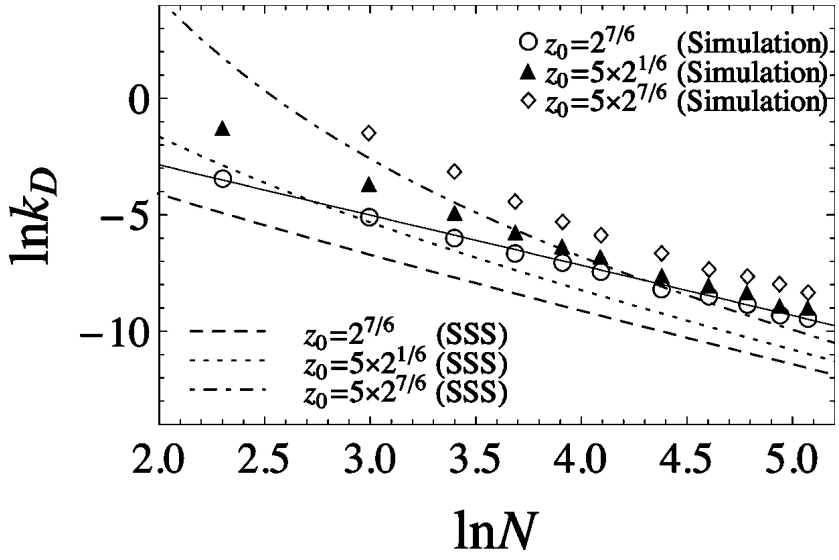


Figure 3.2: In the limit where the chain length is long enough that the typical distance z from its end to the surface is much longer than the contact distance z_0, the diffusion-controlled rate k_D scales as $k_D \propto N^{-b}$ (solid line) where b is close to 2.16. Deviations from this power law are observed when the typical values of z become comparable to z_0. The SSS theory (dashed lines) describes well both the asymptotic power law and departures from this law observed at large z_0 and small N.

is obviously incorrect, as the probability of finding the end monomer at $z = 0$ is identically equal to zero, and so collisions defined by $z_0 = 0$ never happen. In contrast, the SSS rate of Eq. 3.11 exhibits physically reasonable behavior and vanishes at $z_0 = 0$. The regime of small contact distances that are shorter than or comparable to the monomer size is physically quite different from the regime considered here. For $z_0 < \sigma$, the properties of the distribution $p(z)$ are strongly dependent on the interaction potential between the end monomer and the surface. Therefore, the behavior of k_D can be strongly model-dependent. In contrast, results should become less model dependent when the contact

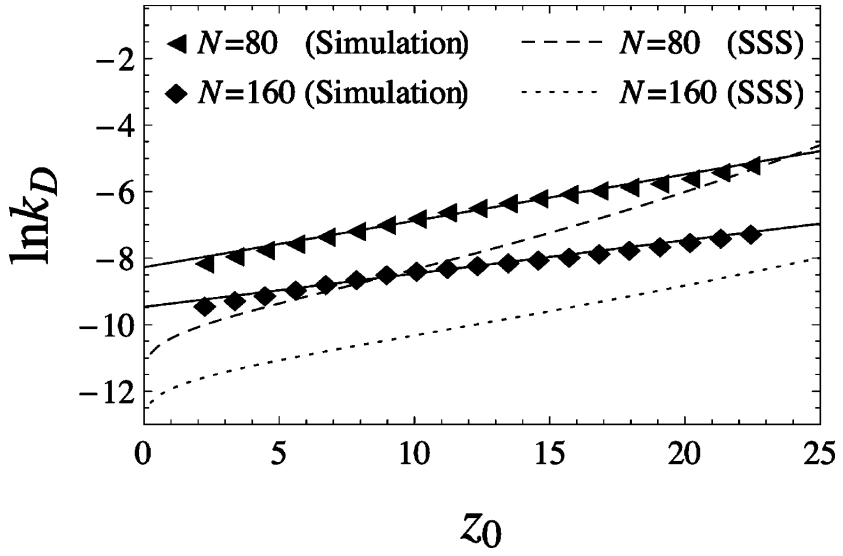


Figure 3.3: The diffusion controlled rate k_D increases (approximately exponentially) with the contact distance z_0. SSS predicts similar behavior, although the functional dependence on z_0 is not exponential Eq. 3.11.

distance z_0 exceeds a typical length scale of surface-monomer interactions. Below we propose that, for $z_0 \gg \sigma$ and $N \gg 1$, the dependence $k_D(N, z_0)$ can indeed be written in a simple, universal form.

3.4.2 Dimensional analysis of $k_D(N, z_0)$

Let us first ignore excluded volume interactions among the non-bonded beads of the chain. Assuming the overdamped limit (i.e., neglecting the first term in Eq. 3.4) and taking the long chain limit ($N \gg 1$), our model becomes equivalent to the standard Rouse model [44]. Treating the monomer number n as continuous, relaxation of the chain is then described by the equation [43,44]

$$\frac{\partial \mathbf{r}_n}{\partial t} = \frac{\pi^2 N^2}{t_R} \frac{\partial^2 \mathbf{r}_n}{\partial n^2} \tag{3.19}$$

where $\mathbf{r}_n$ is the position of the n-th monomer and t_R is the Rouse time defined in Eq. 3.16. For a chain with one end fixed, the boundary conditions are: $\mathbf{r}_0 = 0$ (monomer does not move) and $(\partial \mathbf{r})/(\partial n) \mid_{n=N} = 0$ (zero force on the free end). This leads to the solutions of the form $\mathbf{r}_n(t) \sim \exp(-t/t_p) \sin[\pi(1/2 + p)n/N]$ for $p = 0, 1, 2, ...$, where the relaxation times are given by $t_p = t_R/(1/2 + p)^2$. The lowest relaxation time is thus $t_0 = 4t_R$, as already pointed out in Section 3.3.3.

In the reduced units $\zeta = n/N$, $\mathrm{T} = t/t_R$, and $\rho(\zeta) = \mathbf{r}_n/\langle z^2 \rangle$, Eq. 3.19 becomes invariant with respect to chain length:

$$\frac{\partial \rho}{\partial \mathrm{T}} = \frac{1}{\pi^2} \frac{\partial^2 \rho}{\partial \zeta^2} \tag{3.20}$$

Likewise, when viewed at a spatial resolution that is coarser than the monomer length scale σ, the spatial properties of the chain (in particular the spatial distribution of z) are invariant when the distances are normalized by the only remaining characteristic length scale $\langle z^2 \rangle^{1/2}$(see, e.g., refs [43] and [76]). We thus expect that for $z_0 \gg \sigma$, the first passage time τ when scaled by t_R should be a dimensionless quantity that only depends on the dimensionless ratio $u = z_0/\langle z^2 \rangle^{1/2}$, i.e.,

$$\tau/t_0 = \tau/(4\tau_R) = 1/f(u) \tag{3.21}$$

where $f(u)$ is some yet unknown function.

The SSS approximation for the diffusion-controlled rate is precisely of the form of Eq. 3.21. Indeed, using Eq. 3.17, Eq. 3.11 can be rewritten in the

form of Eq. 3.21, where

$$f(u) = 1/\Gamma(0, u^2) \tag{3.22}$$

We now surmise that Eq. 3.21 should be valid even in the more general case where the polymer is not Gaussian, provided that a characteristic polymer reconfiguration time scale t_0 is appropriately defined. Specifically, we can estimate t_0 by fitting the autocorrelation function of Eq. 3.14 by an exponential (cf. Eq. 3.15).

If our hypothesis is correct, all of the data shown in Fig. 3.2 should collapse onto a single curve when plotted as $k_D t_0$ versus $u = z_0/\langle z^2 \rangle^{1/2}$. Replotting Fig. 3.2 this way, we do not find a perfect collapse (Fig. 3.4). Nevertheless, deviations from a hypothetical single curve observed in Fig. 3.4 are less than a factor of 2, while the rate k_D itself spans many orders of magnitude (cf. Fig. 3.2). It is likely that the lack of perfect collapse can be attributed to the fact that the asymptotic $N \to \infty$ behavior has not been attained for our chains; indeed, it has been previously observed [74] that finite-size effects may persist in the statistics of spatially confined polymer chains even when the latter are hundreds of times longer than the Kuhn segment. An ultimate resolution of this issue would be to perform simulations with much longer chains but, unfortunately, it is impractical to obtain statistically accurate data for such long chains given our central processing unit (CPU) resources. Given these limitations, nevertheless, it is important to point out that for sufficiently short contact distances relative to the typical size of the polymer chain (i.e., for $u \leq 1$), the SSS-derived universal curve predicted by Eq. 3.22 describes all

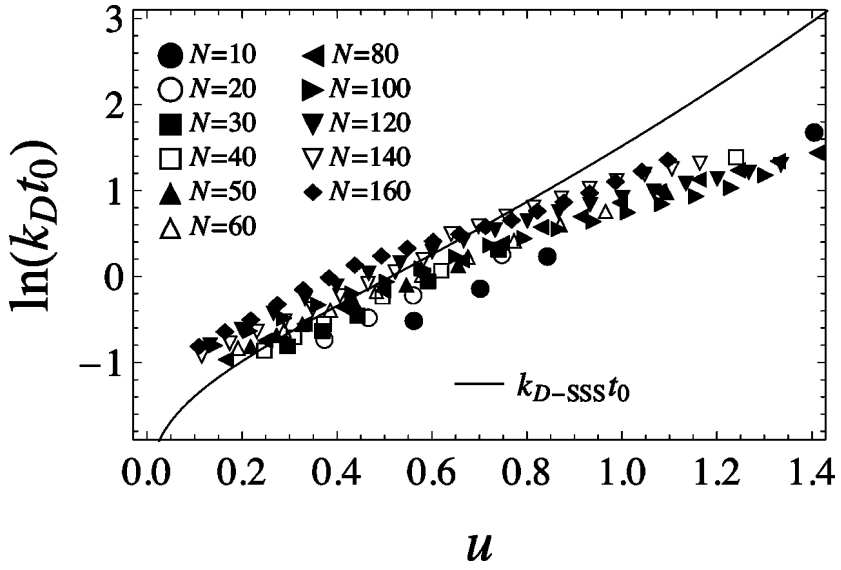


Figure 3.4: For sufficiently long chains, we expect the rescaled rate $k_D t_0$ to asymptotically approach a universal relation when plotted as a function of the dimensionless parameter $u = z_0/\langle z^2 \rangle^{1/2}$. Here we test this prediction and find that this collapse is not perfect for the chains of the lengths as long as $N = 160$ monomers studied here, presumably because of finite-size effects. The solid line represents a plot of $k_{DSSS}t_0 = 1/\Gamma(0, u^2)$ predicted by the SSS theory (Eq. 3.11).

of our data to within a factor of 2.

These findings allow us now to better understand the origins of the power law for the rate k_D as a function of chain length N. Consider the limit where the contact distance z_0 is much shorter than the characteristic polymer size $\langle z^2 \rangle^{1/2}$ (i.e., when $u \ll 1$) but still longer than the monomer size σ. The function $f(u)$ is only weakly dependent on u in this limit. For example, the SSS prediction for $f(u)$ (Eq. 3.22) gives logarithmic dependence in this regime:

$$f(u) \approx -(2 \ln u + \gamma)^{-1} \tag{3.23}$$

where $\gamma = 0.577$ is Eulers constant. If we neglect this weak dependence, then, according to Eq. 3.21, the diffusion-controlled rate should scale as $1/t_R$. For a Rouse chain with excluded volume, the scaling of t_R has been predicted [76] to be of the form $t_R \propto \langle R^2 \rangle / D \propto N^{1+2\nu} \approx N^{2.2}$, where $\nu \approx 3/5$ is Flory's scaling exponent [43], R is the end-to-end distance of the polymer, and D is the diffusion coefficient for the entire chain. Thus we expect the scaling of the form $k_D \propto N^{-2.2}$ for excluded-volume polymer chains. The actual scaling exponent estimated numerically (Fig. 3.2) is indeed close to this prediction. When u is no longer small, then the chain length dependence of $f(u)$ can no longer be neglected. This leads to a stronger chain length dependence of the overall rate constant k_D and to deviations from the simple power law of Eq. 3.18 (Fig. 3.2).

We finally note that similar $1/t_R$ scaling has been predicted for the related problem of the diffusion-controlled rate of end-to-end collisions in a freely diffusing chain [20, 21, 24, 28]. We will further compare these two cases in Section 3.6.

3.5 Results: The reaction-controlled rate k_R

Similarly to the diffusion controlled rate, the reaction-controlled rate k_R (Eq. 3.8) exhibits a power law behavior, $k_R \propto N^{-c}$, with the exponent c close to 1, provided that z_0 is much smaller than the typical end-to-surface distances z (Fig. 3.5). This scaling exponent is also close to that estimated for a Gaussian chain. This can be seen by substituting the distribution $p(z)$ from

Eq. 3.10 into Eq. 3.8, which gives

$$\frac{k_R}{k_0} = 1 - \exp(-u^2), u = \frac{z_0}{\langle z^2 \rangle^{1/2}} \tag{3.24}$$

For $z_0 \ll \langle z^2 \rangle^{1/2}$, this equation predicts $k_R/k_0 \approx u^2 = z_0^2/\langle z^2 \rangle \propto 1/N$ since $\langle z^2 \rangle \propto N$ for a Gaussian chain. Crudely speaking, k_R is proportional to the probability $p(z_0)$ of finding the chains end within the required distance from the surface, which, according to Eq. 3.10, is proportional to $1/N$ for Gaussian chains. Curiously, our data indicates that the approximate $1/N$ scaling is rather robust and insensitive to the chain statistics (Fig. 3.5). Moreover, the same scaling is found in the opposite extreme in which the chain is very stiff and behaves like a rigid rod of length L. Indeed, for a rigid rod that is free to pivot around its attachment to the surface, the probability distribution $p(z)$ is given by $p(z) = (2/\pi L)(1 - z^2/L^2)^{1/2} \approx 2/(\pi L)$. Thus the rate k_R scales again as $1/L \propto 1/N$.

Regardless of whether excluded volume effects are present, the reaction controlled rate for sufficiently long chains satisfies the scaling relationship

$$k_R/k_0 = g(u) = g(z_0/\langle z^2 \rangle^{1/2}) \tag{3.25}$$

provided that z_0 is greater than the monomer size. This scaling relationship immediately follows from Eq. 3.8 and from the fact that $p(z)$ is of the general form [43, 76] $p(z) = \langle z^2 \rangle^{-1/2} h(z/\langle z^2 \rangle^{1/2})$, where $h(x)$ is another dimensionless function.

Consistent with the above prediction, the data of Fig. 3.5, when re-plotted as k_R/k_0 versus u, collapses onto a universal curve $g(u)$ (Fig. 3.6).

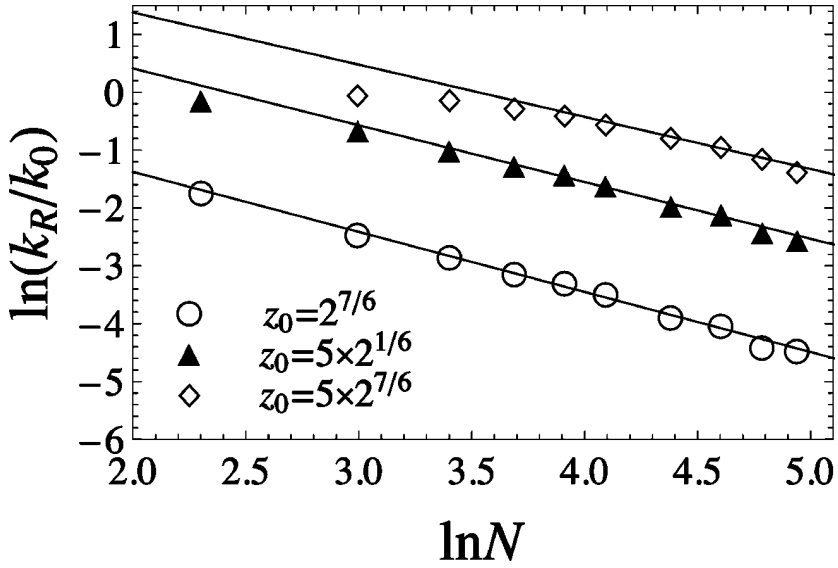


Figure 3.5: Chain length dependence of the reaction-controlled rate k_R is close to inversely proportional. The solid lines are power law fits N^{-b}, where b =1.04, 0.981, and 0.901 for $z_0 = 2^{7/6}\sigma$, $z_0 = 5 \times 2^{1/6}\sigma$, and $z_0 = 5 \times 2^{7/6}\sigma$, respectively.

Moreover, the relationship $g(u) = 1 - \exp(-u^2)$ predicted for ideal chains (Eq. 3.24) describes the data reasonably well even in the presence of excluded volume effects.

3.6 End-to-surface versus end-to-end reaction

Although collisions of the free end of a surface-attached DNA with the respective surface are believed to provide the mechanism through which target-induced structural changes are detected in E-DNA sensors (see, e.g., refs [8] and [7]), very little experimental information is currently available about the basic physics of this process. In contrast, collisions between the ends of an un-

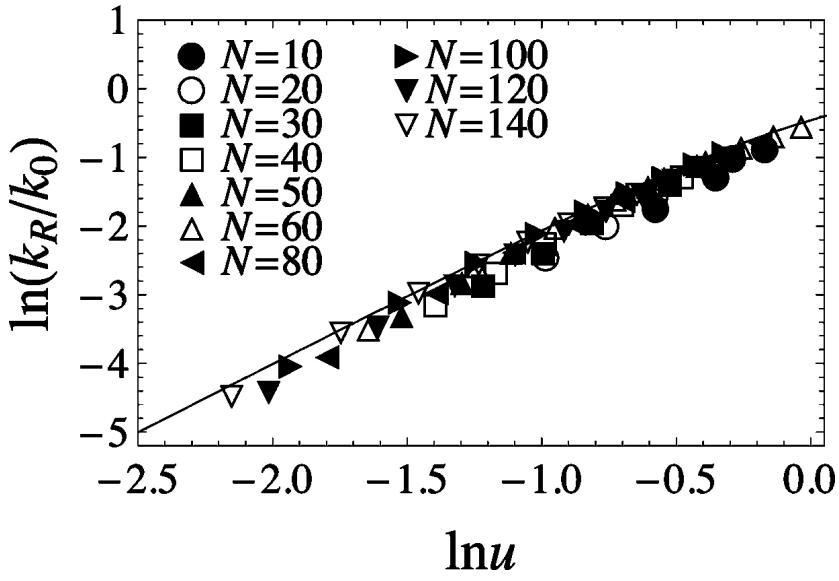


Figure 3.6: k_R/k_0 approaches a universal relation when plotted as a function of the dimensionless parameter $u = z_0/\langle z^2 \rangle^{1/2}$. The solid line is the analytical prediction for Gaussian chains (Eq. 3.24)

constrained biopolymer (ssDNA or protein) in solution have received considerable experimental attention [2,22,34,35,37,41]. It is therefore natural to ask whether the current understanding of end-to-end reaction dynamics (Fig. 3.1) is transferrable to the problem of an end-to-surface reaction (Fig. 3.1). To answer this question, here we compare the simulation results reported in the previous sections with end-to-end reaction rates for the same polymer chains. For the latter process, we have adopted a model analogous to that described in Section 3.3.2. Specifically, we have assumed that a chain-mediated end-to-end reaction has an end-to-end distance-dependent rate described by Eq. 3.5, where z is replaced by the distance R between the end monomers of the polymer (Fig. 3.1). The parameter z_0 is now interpreted as a capture radius such

that no reaction occurs if the end-to-end distance exceeds it. The results for end-to-end rates reported here were obtained through a straightforward modification of the procedures described in Section 3.3: Eqs. 3.5, 3.6, 3.7, and 3.8 were used to compute the diffusion- and reaction-controlled rates by simply removing the surface constraint and replacing z by R.

The scaling of the diffusion-controlled rate with chain length N is very similar in the cases of end-to-end collisions in unconstrained polymers and of end-to-surface collisions of surface-attached polymers (Fig. 3.7). This result is not surprising: According to earlier theoretical predictions for unconstrained polymers [20,21,28] and to the scaling considerations of Section 3.4, k_D should roughly scale as $1/t_R$ in each case, resulting in a similar chain length dependence (Eq. 3.18) with a scaling exponent close to $b = 2.2$. For sufficiently small z_0, we find that the end-to-surface rate is always greater than the end-to-end rate. Again, this result is not surprising since a surface presents a larger target for a chain end to strike, as compared to the other end monomer of the chain.

Just as in the case of collisions with a surface, the end-to-end diffusion controlled rates measured for chains of varying length and capture radius z_0 fall close to a universal curve when properly rescaled. Specifically, we find a relationship of the form (Fig. 3.7)

$$k_D t_0 = \widetilde{f}(u) \tag{3.26}$$

Similarly to the case of surface-attached chains, the parameter u for an un-

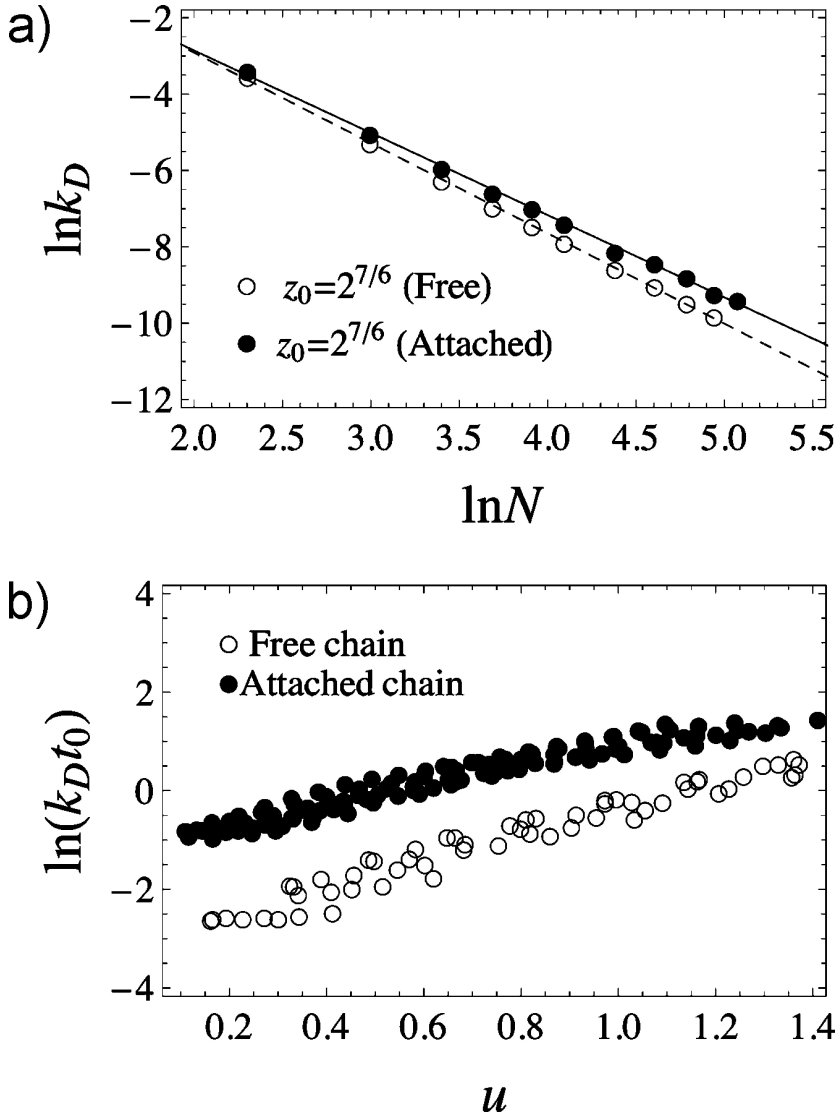


Figure 3.7: (a) End-to-end collisions of a free chain and end-to-surface collisions of a tethered chain show similar chain length dependence in the diffusion controlled regime. Here, $k_D \propto N^{-2.37}$ and $N^{-2.16}$ in the former and latter cases for $z_0 = 2^{7/6}\sigma$. (b). The diffusion-controlled rate of end-to-end collisions for a free chain and end-to-surface collisions of a tethered chain exhibit similar (near) universal dependence on the reduced parameter $u = z_0/\langle z^2 \rangle^{1/2}$ when measured in reduced time units set by the polymer relaxation time scale t_0. The data for surface attached chains are the same as in Fig. 3.4

constrained chain is defined as the ratio of z_0 to the rms chain extension in the z direction (which is now equivalent to the x- and y-directions)

$$u^2 = z_0^2/\langle z^2 \rangle = 3z_0^2/\langle R^2 \rangle \tag{3.27}$$

The relaxation time t_0 was determined by fitting the autocorrelation function of the end-to-end distance to an exponential, similarly to Eqs. 3.14 and 3.15. The function $\widetilde{f}(u)$ defined this way behaves very similarly to the function $f(u)$ found for end-to-surface collisions (Fig. 3.7). For the same values of t_0 and u, the rate of end-to-surface collisions is roughly an order of magnitude higher than that of end-to-end collisions (Fig. 3.7).

Similarly to the case of diffusion-controlled rates, the reaction-controlled rate k_R for an end-to-end reaction is smaller than that for an end-to-surface reaction (Fig. 3.8). In contrast to the diffusion-controlled rate, however, the scaling of the reaction-controlled rate as a function of chain length is quite different for end-to-end and end-to-surface reactions. This difference in scaling is readily understood by considering the case of a Gaussian chain. For a surface-attached Gaussian chain, the rate k_R is given by Eq. 3.24. For long chains this rate thus scales as $k_R \propto u^2 \propto N^{-1}$. In contrast, the reaction controlled end-to-end rate for a Gaussian chain is given by (cf. Eq. 3.8)

$$k_R/k_0 = \int_0^{z_0} p(r)\,\mathrm{d}r = \mathrm{erf}\left(\frac{u}{\sqrt{2}}\right) - \left(\frac{2}{\pi}\right)^{1/2} u \exp\left(-\frac{u^2}{2}\right) \tag{3.28}$$

For $u \ll 1$, this results in chain length dependence given by $k_R \propto u^3 \propto N^{-3/2}$. These scaling laws are easy to understand: As noted in Section 3.5,

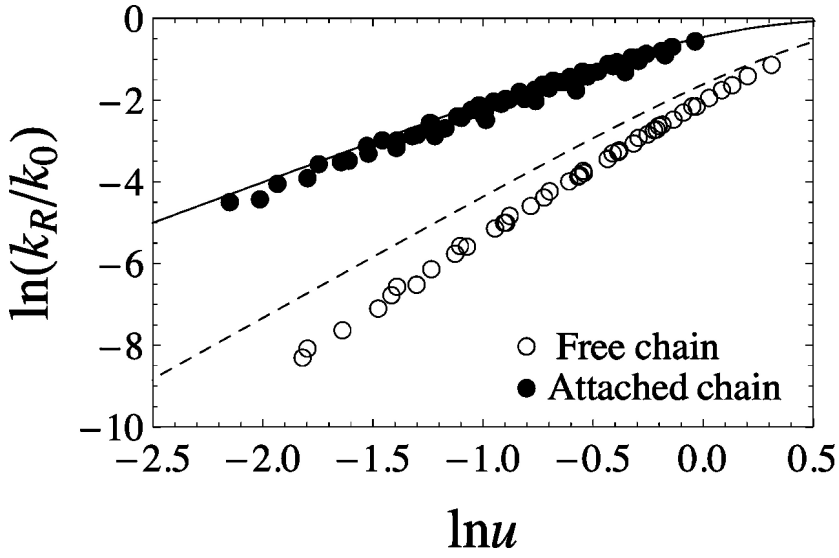


Figure 3.8: In the reaction controlled regime, the rate of an end-to-end reaction for an unconstrained chain has a stronger chain length dependence than that of an end-to-surface reaction of a surface tethered chain, according to both numerical simulations (circles) and analytical theory for Gaussian chains (lines). Here, these rates are plotted as a function of the dimensionless parameter $u = z_0/\langle z^2 \rangle^{1/2}$, which itself scales as $u \propto N^{-\nu}$ (where $\nu = 1/2$ and $3/5$ for Gaussian and excluded-volume chains, respectively).

the reaction-controlled rate is proportional to the probability of finding the reacting entities in close proximity. This probability scales as $1/N$ for a chain end reacting with the surface and as $N^{-3/2}$ for one chain end reacting with the other. Consequently, an end-to-end reaction exhibits a significantly stronger dependence on chain length as compared to its end-to-surface analog in the reaction limited regime. This conclusion remains true when excluded volume effects are taken into account (see Fig. 3.8). Interestingly, the Gaussian chain approximation works reasonably well in the case of a surface-attached chain (Eq. 3.24), while it is considerably less accurate in the case of an end-to-end

reaction within an unconstrained polymer.

3.7 Concluding Remarks

Our study was largely motivated by the recently developed E-DNA method of detecting ssDNA [7, 8]. In this technique, an ssDNA probe attached to an electrode surface undergoes hybridization with a free DNA thus resulting in the formation of a dsDNA tethered to the surface. This changes the electron transfer rate between the redox-modified free DNA end and the surface, which is manifested as a change in the Faradic current. To understand this change more quantitatively, first note that the Kuhn length l_K of dsDNA is much longer than that of ssDNA [77]. Since, depending on the intrinsic rate of electron transfer, both reaction-limited and diffusion-limited regimes are possible, below we consider both cases separately.

Diffusion-Controlled Case. Using Eq. 3.21 and Eq. 3.22 and omitting numerical factors, we can write the diffusion-limited rate as

$$k_D \propto t_0^{-1} f(z_0/\langle z^2 \rangle^{1/2}) \propto t_0^{-1} f(u) \tag{3.29}$$

The characteristic chain relaxation time can often be approximated by the following scaling relationship [43] (Eq. 3.16 being its particular case for the Rouse model):

$$t_0 \propto \langle R^2 \rangle / D_{chain} \tag{3.30}$$

where $\langle R^2 \rangle$ is the mean square end-to-end distance and D_{chain} is the diffusion constant of the entire chain. Using $\langle R^2 \rangle \sim \sigma l_K N$ we find that the factor

t_0^{-1} in Eq. 3.29 is roughly inversely proportional to the Kuhn length and thus should decrease upon DNA hybridization. In addition, it is plausible that the diffusion coefficient Dchain for a dsDNA is lower than that of an ssDNA of the same length, thus providing further hybridization-induced reduction of k_D. Finally, since $f(u)$ is a monotonically increasing function of the parameter

$$u = z_0/\langle z^2\rangle^{1/2} \propto z_0/(N\sigma l_K)^{1/2} \tag{3.31}$$

the last factor in Eq. 3.29 will also decrease upon DNA hybridization. All these effects will result in a reduction of the Faradic current, consistent with the experimental observations [7, 8].

Reaction-Limited Case. In this limit, we can use Eq. 3.24 to estimate the dependence of the rate k_R on the polymer Kuhn length. If we further assume that $z_0 \ll \langle z^2\rangle^{1/2}$, then we find

$$k_R \propto \langle z^2\rangle^{-1} \propto l_K^{-1} \tag{3.32}$$

Like in the diffusion-controlled case, the reaction-controlled rate decreases with increasing Kuhn length, but this dependence is somewhat weaker than for the diffusion-limited rate.

The Limit of Stiff dsDNA. In the above estimates, we assumed that the polymer contour length $L = N\sigma$ is much longer than its Kuhn length so that both the ssDNA and the dsDNA are flexible. However, while the Kuhn length of ssDNA typically corresponds to just a few bases, the Kuhn length of dsDNA corresponds to over 100 base pairs, which is longer than

the length of the DNA constructs commonly employed in DNA sensors. For such short and therefore stiff DNA, the end-to-surface collision rate strongly depends on the manner in which the molecule is attached to the surface. For example, if the molecule is not free to pivot around the attachment point, then a collision of its end with the surface will involve surmounting an energy barrier required to bend the molecule. This scenario has been studied in ref 16. Here we consider the opposite case of a dsDNA approximated as a rigid rod freely pivoting around its surface attachment. The diffusion limited rate can be crudely estimated using a relationship similar to Eq. 3.30, as the inverse time it takes the whole chain to diffuse over the distance comparable to its own size:

$$k_{D,dsDNA} \sim \frac{D_{chain,dsDNA}}{L^2} \tag{3.33}$$

where $L = N\sigma$ is the length of the molecule. To compare this result with the end-to-surface collision rate for ssDNA of the same contour length, we can use Eqs. 3.29 and 3.30. If we further neglect the weak dependence of $f(u)$ and replace it by a constant, we can write

$$k_{D,ssDNA} \sim \frac{D_{chain,ssDNA}}{\langle R^2 \rangle} = \frac{D_{chain,ssDNA}}{N\sigma l_{K,ssDNA}} \tag{3.34}$$

Comparing Eq. 3.33 and Eq. 3.34, we find that $k_{D,ssDNA} \gg k_{D,dsDNA}$ as long as the contour length of the DNA is longer than the Kuhn length of ssDNA, $L = N\sigma \gg l_{K,ssDNA}$.

In the reaction-controlled limit, we use the fact that the probability that the free end of a rigid rod is within a distance $z_0 \ll L$ from the surface is

given by $\sim p(0)z0 \approx 2z_0/(\pi L)$ (see Section 3.3). Thus the reaction-limited rate is given by $k_{R,dsDNA} \propto k_0 z_0/L$, which, again, should be compared with the ssDNA rate estimated from Eq. 3.24, $k_{R,ssDNA} = k_0 z_0^2/(N\sigma l_{K,ssDNA})$. Comparing the two rates, we find that $k_{R,ssDNA} \gg k_{R,dsDNA}$ as long as $z_0 \gg l_{K,ssDNA}$. The latter inequality, however, is not necessarily satisfied under experimental conditions. Indeed, if the process in question involves electron transfer, the corresponding characteristic length scale may be very short, on the order of a few angstroms shorter than the typical Kuhn length of ssDNA. The case $z_0 < l_{K,ssDNA}$ would have to be dealt with much more carefully. In particular, Eq. 3.24 is likely to be inadequate in this case because the probability distribution $p(z)$ for such short distances would be affected by the specifics of the polymer-surface interaction. Finally, since the dynamics of dsDNA is effectively slower than that of ssDNA, it is also conceivable that a transition from reaction-controlled to diffusion-controlled electron transfer kinetics can happen upon DNA hybridization. Further understanding of how DNA hybridization affects the electron transfer rate in E-DNA sensors will require more realistic polymer models as well as knowledge of the specifics of the electron transfer system employed (such as the intrinsic quenching rate and its distance dependence).

Chapter 4

Universality in the timescales of internal loop formation in unfolded proteins and single-stranded oligonucleotides

4.1 Abstract

Understanding the rate at which various parts of a molecular chain come together to facilitate the folding of a biopolymer (e.g., a protein or RNA) into its functional form remains an elusive goal. Here we use experiments, simulations, and theory to study the kinetics of internal loop closure in disordered biopolymers such as single-stranded oligonucleotides and unfolded proteins. We present theoretical arguments and computer simulation data to show that the relationship between the timescale of internal loop formation and the positions of the monomers enclosing the loop can be recast in a form of a universal master dependence. We also perform experimental measurements of the loop closure times of single-stranded oligonucleotides and show that both these and previously reported internal loop closure kinetics of unfolded proteins are well described by this theoretically predicted dependence. Finally, we propose that experimental deviations from the master dependence can then be used as a sensitive probe of dynamical and structural order in unfolded proteins and other biopolymers.

4.2 Introduction

Conformations adopted by disordered polymers and the timescales of interconversion among them are thought to play key roles in biomolecular folding [2–5, 78] and function [79]. For example, the discovery that the folding rates of single-domain proteins are reasonably strongly correlated with various measures of their native state topology [80, 81] has been interpreted in terms of the efficiency with which the largely unfolded polypeptide chain undergoes loop closure and other polymer rearrangements as it stochastically searches for the correct, native-like overall topology [5, 33].

Likewise, the nature of the disordered state of biopolymers, including both polypeptides (reviewed in Oh et al. [82]) and single-stranded nucleic acids (reviewed in Lubin and Plaxco [83]) has further importance in the context of recent biosensor designs, which rely on target-induced changes in the dynamics and/or conformational ensembles of their component biopolymer probes [7, 8].

Despite the importance of biomolecular dynamics, current understanding of the structural conformations and intramolecular dynamics in unfolded proteins and single-stranded oligonucleotides remains incomplete. For example, the amount of residual order in chemically unfolded proteins remains a contentious issue [84], with some experimental [85] and computational [86, 87] evidence pointing to a significant degree of structural organization, whereas other measurements [9–11,88,89] exhibit Flory's random-coil scaling of the spatial dimensions with the polypeptide chain length N (number of monomers).

In an effort to better understand the dynamics of unstructured biopolymers, a number of groups have studied the dynamics of end-to-end collisions using both theoretical [3, 13, 20, 22–25, 28, 29, 31, 32, 64, 90–93] and experimental [2, 4, 6, 22, 35, 37, 63, 94] approaches. In contrast, studies of the rate with which two internal positions in a chain collide, or a terminus collides with an internal position, remain limited [34, 69, 95–101]. Here, we explore this specific dynamic property of an unfolded biopolymer via simulations, experiments, and theory.

Experimentally, the frequency of such intrachain collisions can be measured by monitoring the quenching kinetics of a suitably chosen probe, which is attached to one monomer, by a near contact-limited quenching group that is linked to another monomer [2]. In the diffusion-limited regime, where the intrinsic rate of quenching is so high that it occurs virtually instantaneously upon a collision between the monomers, the quenching kinetics is governed by the timescales of polymer motion and provides a measure of the collision frequency. This frequency depends on the location of the chosen pair of monomers within the chain. For a chain of $N+1$ monomers labeled 0,1,..., N, the central quantity considered here is the mean time $\tau_{ij}(N)$ to form a loop of length $|i-j|$ as a result of a collision between the monomers i and j within the chain (Fig. 4.1). End-to-end (EE) collisions then correspond to $i=0$, $j=N$, internal-to-end (IE) collisions correspond to $i=0$, $j<N$ or $j=N$, $i>0$, and internal-to-internal (II) collisions to $0<i,j<N$.

The kinetic effect of the tails−the set of monomers external to the

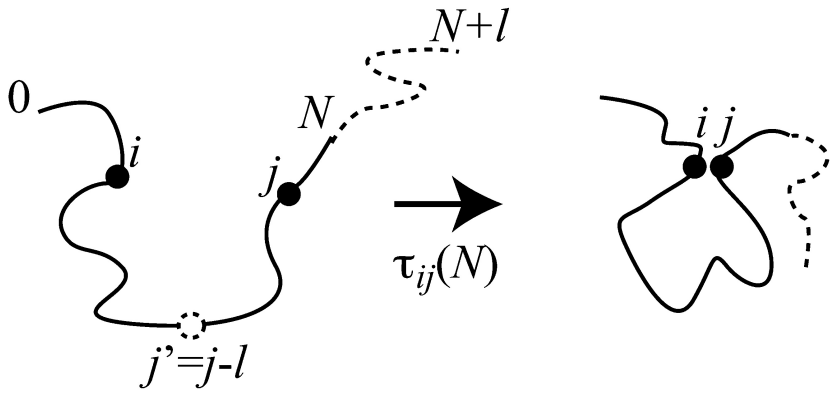


Figure 4.1: Collision between monomers i and j within a polymer chain leads to the formation of a loop of length $|i-j|$. Here we describe how the timescale for this process depends on the monomer positions and the total chain length N. Of particular interest is the effect of tails, i.e., chain segments exterior to the loop. Appending a tail of length l to one of the chain ends (*dashed line*) increases the loop closure time. The tail effect is less obvious in a scenario where one of the monomers is moved into another position (*dashed-line circle*) such that the length of a tail is increased while the total chain length remains constant.

loop−has been previously studied experimentally [34, 95], theoretically [97, 102], and via simulations [64, 98, 99, 102]. In particular, when the loop length $|i-j| = n$ was fixed and the tail length l was increased (Fig. 4.1), the IE collision time $\tau_{0n}(n+l)$ was found to increase monotonically to saturation, achieving the limit of a finite loop within a semiinfinite chain. Qualitatively, this result is easy to understand: more monomers have to move in order to close a loop whenever a tail is present [44]. In addition, tails introduce steric clashes that reduce the probability and, consequently, slow the rate of loop closure [95, 99, 103]. Although the experimental studies of IE loops used tails of finite length, the theoretical limit of infinite tails has also been studied,

leading to scaling relationships between the loop formation time and its length [97]. Finally, effects of chain flexibility on the interior loop formation within semiflexible chains have also been studied theoretically [102].

Although the above studies have provided initial forays into the dynamics of internal loop closure, current understanding of internal loop dynamics in biopolymers remains far from complete.

First, unfolded proteins and DNA typically used in experimental studies are far from an asymptotic infinite-chain limit that is commonly assumed by polymer theories. Finite-size effects, i.e., deviations from the behavior expected as $N \to \infty$, as well as sequence-specific effects, thus may significantly affect the experimentally observed loop formation dynamics.

Second, even if the effects of the sequence and of finite chain length are neglected a number of theoretical issues remain to be resolved. To illustrate those, imagine two different experiments where the length of the tail adjacent to the monomer j is increased by l (Fig. 1). In the first one, the loop length remains constant while a chain segment of length l is appended to the polymer. As a result, $\tau_{ij}(N + l)$ increases with the tail length l, as established by the abovementioned studies. In the second experiment, however, the monomer j is moved into a new position $j' = j - l$ while the total chain length remains constant.

Will the loop time $\tau_{i,j-l}(N)$ increase or decrease with increasing l?

The answer to this question involves a trade-off that has seen little

previous exploration: as the tail grows longer, which will slow dynamics, the loop grows concomitantly shorter, which will speed it up and prior theory makes no prediction as to which of these two effects wins. In what follows we report on experiments−both in vitro and in silico−that explore the second of the above two scenarios, wherein the locations of the colliding entities are varied within a polymer chain of a given length. We will show that the resulting dependence of the loop formation time on the tail length may be nonmonotonic such that the loop formation time increases at first and then decreases with l.

More importantly, we will demonstrate the existence of a universal relationship between the loop formation timescale, the fundamental time- and length-scales of the entire polymer chain, and the relative positions of the colliding entities within the chain. Moreover, we will show that this observation is universal not only in a conventional, polymer-theoretical sense, but also applies, with a surprisingly good accuracy, to finite, experimentally addressable systems. That is, this relationship not only holds theoretically, in the asymptotic, infinitely long chain limit for polymers in a single universality class (e.g., Gaussian chains, excluded volume random flight chains, etc.), but is also seen experimentally across a diverse set of real biopolymers. Finally, we will argue that significant deviations from our theoretical predictions can be used as a sensitive probe to detect signatures of residual order in biopolymers (e.g., unfolded proteins).

4.3 Methods

4.3.1 Simulation studies of loop formation in polymers

Our model for a polymer chain consists of $N + 1$ beads of mass m connected by N springs. Consecutive (bonded) beads separated by a distance r are connected y a harmonic interaction potential,

$$V_{bond}(r) = \tfrac{1}{2} k_{bond}(r - \sigma)^2$$

where σ is the equilibrium bond length and $k_{bond} = 100\epsilon/\sigma^2$. Excluded volume interactions are represented by a repulsive Lennard-Jones potential between nonconsecutive (nonbonded) beads separated by a distance r:

$$V_{nonbonded}(r) = 4\epsilon[(\sigma/r)^{1/2} - (\sigma/r)^6]\theta(2^{1/6}\sigma - r)$$

Here ϵ is a characteristic energy scale and θ is the Heaviside step function that truncates the attractive portion of the Lennard-Jones potential. The dynamics of each bead is governed by the Langevin equation

$$m\ddot{\mathbf{r}}_i = -\tfrac{\partial V}{\partial \mathbf{r}_i} - \xi\dot{\mathbf{r}}_i + \mathbf{F}_i(t)$$

where $\mathbf{r}_i(t)$ is the position of the bead, V is the total interaction potential, xi is a friction coefficient, and $\mathbf{F}_i(t)$ is a random force satisfying the fluctuation-dissipation theorem. A friction coefficient of $\xi = 2(\sigma^2/m\epsilon)^{-1/2}$ was chosen such that the dynamics were in the overdamped regime [12, 14]. For the case where hydrodynamic interactions were included, we used the Ermak-McCammon

Brownian dynamics algorithm [104] with the Rotne-Prager-Yamakawa diffusion tensor [105]. The hydrodynamic radius for each monomer bead was set at 0.35σ. All simulations were performed at a temperature of $1\epsilon/k_B$. The results reported here are given in dimensionless units, where $\tau_0 = (m\sigma^2/\epsilon)^{1/2}$ sets the unit of time.

To compute the loop formation time between a selected pair of monomers, i and j, we have assumed that a loop is formed instantaneously whenever the monomers are within a distance R_c of one another. In practice, we introduce a distance-dependent rate [12] given by

$$k_{ij}(r_{ij}) = k_0\theta(R_c - r_{ij})$$

where r_{ij} is the distance between monomers i and j and k_0 is the intrinsic rate, which is chosen to be large enough that the diffusion-controlled limit is reached and the results of the simulations are independent of k_0.

The mean time τ_{ij} for diffusion-controlled collisions between a pair of monomers i and j is computed from the equation

$$\tau_{ij} = \int_0^\infty S_{ij}(t)\,\mathrm{d}t$$

where

$$S_{ij}(t) = \langle \exp(-\int_0^t k_{ij}[r_{ij}(t')]\,\mathrm{d}t') \times \theta[r_{ij}(0) - R_c]\rangle / \langle\theta[r_{ij} - R_c]\rangle$$

is the probability that no collisions between i and j took place from time 0 to t provided that the initial distance between the two monomers, $r_{ij}(t = 0)$

exceeds R_c. Here the angular brackets denote averaging over the canonical ensemble of initial polymer configurations. See Cheng and Makarov [14] for further computational details.

4.4 Results and Discussion

4.4.1 End-to-end versus internal-to-end loops: Experiment

In a previous study [37], Uzawa et. al measured the diffusion-controlled end-to-end loop formation kinetics in unstructured, single-stranded DNAs as a function of chain length. Here, we report upon an extension of these studies involving the measured internal-to-end loop formation kinetics of single-stranded DNA consisting of 27 monomers (Fig. 4.2). In doing so we find that, in agreement with both previous experiments [34, 95] and simulations [64, 98, 99], the IE loop formation time $\tau_{ij}(N) = \tau_{0j}(N)$ is always longer than the end-to-end collision time $\tau_{0j}(j)$ for a loop of the same length j (Fig. 4.2). At the same time, for chains of a given total length N, the loop formation time decreases as the length of the tail, $N-j$, increases, a trend opposite that observed in the constant loop-length measurements. We also find that, except for the shortest loops, the experimental loop length dependence of the loop formation time can be fitted by a power law for both the EE and the IE cases,

$$\tau_{ij} \propto |i-j|^{\delta} \tag{4.1}$$

with exponents of $\delta = 3.34 \pm 0.1$ and $\delta = 2.83 \pm 0.5$, respectively. We emphasize that Eq. 4.1 should be regarded here as a fit of experimental data rather than

a true scaling law. Indeed, the experimental value of the exponent δ in the EE case is considerably larger than the value predicted by polymer theory in the asymptotic limit of very long chains [97]. As discussed in an earlier article [37], this disagreement between polymer theory and the experimentally observed loop length dependence is likely to arise due to finite-size effects. Specifically, the internal dynamics of the single-stranded DNA constructs that were considered are affected by both electrostatic interactions within these relatively short chains and differences between the DNA and the linkers that connect it to the lumophore and quencher [37].

4.4.2 End-to-end versus internal-to-end loops: Simulations

In addition to the above experiments we have performed Langevin dynamics simulations of loop formation within bead-and-spring polymer models, using the usual assumption (see, e.g., Toan et al. [28] and references therein) that a loop forms between monomers i and j whenever the distance between the two becomes shorter than a certain capture radius R_c. Note that more general, distant-dependent reaction rates may lead to deviations from diffusion-controlled kinetics [12], which is neglected in this study. Our simulations of EE and IE loops show trends similar to those observed experimentally (Fig. 4.3). When the tail is longer than $\sim$35% of the loop length, the loop formation times become independent of the length of the tail. Such saturation in the tail length dependence has been anticipated by theory [97], observed in previous simulations [69, 98] and demonstrated in a recent experimental study [96].

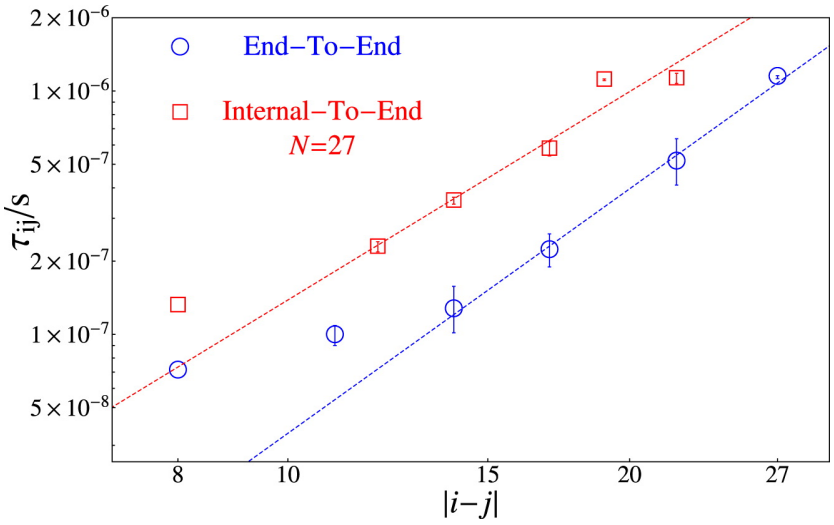


Figure 4.2: Loop length dependence of the loop formation times in single-stranded DNA measured by T. Uzawa and K. W. Plaxco [15]. Except for the shortest loops, the loop length dependence of both internal-to-end (IE) and end-to-end (EE) collision times can be fitted by a power law of the form $\tau_{ij} \propto |i-j|^{\delta}$, where $\delta = 3.34 \pm 0.1$ for EE loops and $\delta = 2.83 \pm 0.5$ for IE loops (*straight lines*). The deviations observed for the shortest loops presumably arise because of the linker effects, as previously reported [37]. Given the same loop length, the loop formation time for an IE loop is always longer than that for an EE loop.

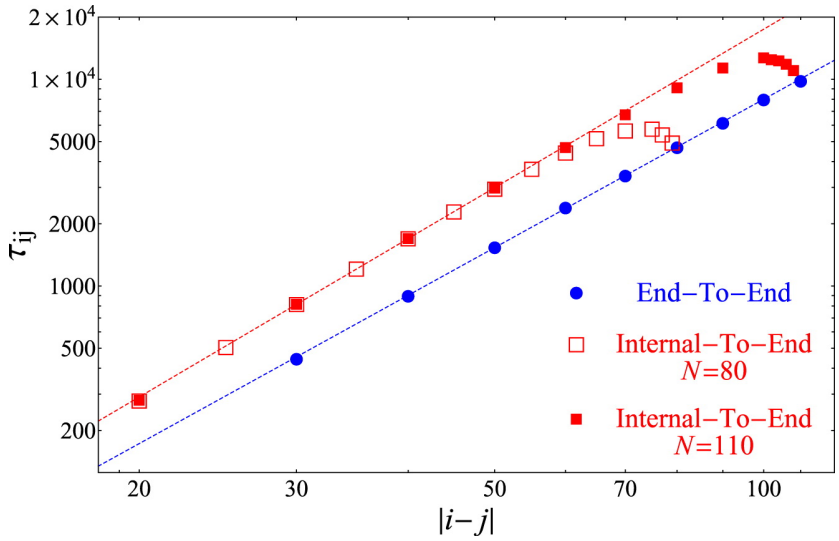


Figure 4.3: Loop length dependence of the loop formation times from simulations of bead-and-spring polymer models. The loop length dependence of both IE and EE collision times can be fitted by a power law of the form $\tau_{ij} \propto |i-j|^{\delta}$, where $\delta = 2.38$ for EE loops and $\delta = 2.55$ for IE loops (*straight lines*). For sufficiently long tails, the IE loop formation time is independent of the tail length (and thus the total chain length). Here, the simulations were performed for Rouse chains with excluded volume interactions. The collision between monomers was assumed to take place whenever the two monomers were within $R_c = 2.5$ equilibrium bond lengths from one another.

In this long-tail limit, the loop length dependence of the IE loop formation time follows a power law of the form of Eq. 4.1. Curiously, unlike the case of EE loops in single-stranded DNA, where the experimental value of δ is significantly higher than that for simple bead-and-spring models [37], the experimental scaling exponent for IE loops observed in Fig. 4.2 agrees well with the bead-and-spring value of $\delta \approx 2.55$ (Fig. 4.3).

This agreement, however, has to be taken with a grain of salt given that the properties of relatively short, single-stranded DNAs are known to deviate from those of idealized random-coil models due, for example, to the importance of electrostatic effects over short length-scales [37]. We also note that the scaling exponents observed in Fig. 4.3 for both EE and IE loops are in reasonable agreement with the renormalization group predictions of Friedman and OShaughnessy [97] as well as with earlier simulations [70]. Indeed, in both cases the value of the scaling exponent predicted by renormalization group theory is [70, 97]

$$\delta = 2\nu + 1 \sim 2.2$$

where ν is Flory's scaling exponent. Considering potentially significant finite size effects that are often observed in simulations of such systems [12, 14], this value appears reasonably close to the $\delta = 2.38$ and $\delta = 2.55$ estimated for EE and IE loops, respectively.

The simulated IE loop formation times are longer than the EE times for the loops of the same length, which is in accord with our experimental

data. Assuming IE loops with infinite tails, the ratio

$$\tau_{0j}(\infty)/\tau_{0j}(j)$$

of the IE and EE loop formation times is not constant but weakly dependent of the loop length j, in contrast to a constant ratio of $\sim$4 predicted by an earlier study of position-dependent reconfiguration times in Rouse chains [106]. Although our simulation results are qualitatively consistent with the experimental data of Fig. 4.2, they also predict a maximum in the dependence of the time $\tau_{0j}(N)$ on j (Fig. 4.3), which is not observed experimentally (Fig. 4.2). This discrepancy is further discussed below.

4.4.3 Dimensional analysis of intrachain loop formation kinetics

Although the qualitative agreement we observe between experiments and simulations is reassuring, development of a more quantitative model for intrachain loop formation is hindered by the fact that experimental loop formation rates depend on the photophysics of the probes used to measure them as well as on the structural properties of the linkers connecting the probes to the molecules of interest [37]. It is difficult to realistically capture such features within coarse-grained bead-and-spring polymer models.

Instead, in our simulations we have assumed that all these effects can be lumped into an effective capture radius R_c that defines a collision. Although this assumption appears plausible, an estimate for the appropriate numerical value of R_c is not readily available for the specific optical reporting groups

and single-stranded DNA polymers that have been experimentally studied. Fortunately, dimensional arguments and numerical examples presented below establish that simulation results can be rescaled to assume a form that is independent of the chain length and is only weakly dependent on the capture radius, thus allowing a direct, quantitative comparison between the loop formation times in idealized, long homopolymer chains and real biopolymers.

Within the model adopted here, the loop formation time $\tau_{ij}(N)$ depends on the capture radius R_c, the solvent viscosity η, the polymer length N, and the positions of the probes i and j within the polymer chain. Generally, the characteristic timescales of polymer dynamics become longer with increasing chain length. In the Rouse model of chain dynamics, for example, these timescales are proportional to N^2 (see, e.g., Refs. [43, 44]).

To study spatio-temporal correlations within chains of different length N, it is sensible to factor out this chain length dependence so that intrachain loop formation times are measured relative to each chain's own characteristic reconfiguration timescale. One could, for example, normalize $t_{ij}(N)$ by the slowest relaxation time for the chain, such as the Rouse (or Zimm) time in case of chains described by the Rouse (or Zimm) model [44]. This choice is, however, impractical because polymer relaxation times are not directly accessible by experimental measurements of loop formation kinetics. Instead, we choose a related timescale [20, 92, 97], the end-to-end loop formation time $\tau_{0N}(N)$, as a measure of the global reconfiguration timescale for the entire

chain. We thus introduce a dimensionless loop formation time given by

$$T_N(i/N, j/N) = \tau_{ij}(N)/\tau_{0N}(N) \tag{4.2}$$

Unlike $\tau_{ij}(N)$, the quantity defined by Eq. 4.2 is independent of the solvent viscosity because both the numerator and the denominator of Eq. 4.2 are proportional to the viscosity [12] in the diffusion-controlled limit. Nondimensionality of $T_N(i/N, j/N)$ further imposes a restriction on its capture radius dependence. Indeed, if we assume that the only two relevant length-scales of the system are the capture radius R_c and a typical length scale R of the polymer, which can be chosen equal to the root-mean-square (RMS) end-to-end distance of the chain, $\langle(\mathbf{r}_N - \mathbf{r}_0)^2\rangle^{1/2}$, then $T_N(i/N, j/N)$ must depend only on the dimensionless ratio, R_c/R, of these two length-scales.

We finally conjecture that the dimensionless time $T_N(i/N, j/N)$, written as a function of chain length and of the monomer positions rescaled by the total length, does not explicitly depend on N so that its subscript can be dropped, i.e., $T_N(i/N, j/N) = T(i/N, j/N)$. This amounts to the assumption that the scaling of $\tau_{ij}(N)$ with loop length for all self-similar loops (i.e., loops sharing the same values of i/N, j/N, and R_c/R) is the same and coincides with that of $\tau_{0N}(N)$. Such self-similarity has been shown to hold for equilibrium intermonomer distance distributions in excluded-volume polymer chains [97, 107] and for reconfiguration times in Rouse-type chains as determined via fluorescence energy transfer experiments [106].

To summarize the above-dimensional analysis, our conjecture is that

the internal loop formation time normalized by the end-to-end loop formation time,

$$\tau_{ij}(N)/\tau_{end-to-end}(N) = T\left(\frac{i}{N}, \frac{j}{N}, \frac{R_c}{R}\right) \tag{4.3}$$

is independent of the polymer length and is a universal function of the length of each tail relative to the total polymer length and of the capture radius normalized by the RMS polymer end-to-end distance. Moreover, although both internal and end-to-end loop formation times may significantly depend on the capture radius, it seems probable that such dependence would at least partially cancel out in a ratio of the two timescales resulting in a weak, or even nonexistent dependence on capture radius.

To test the above conjecture we have computed normalized loop formation times for chains of different length N. When the value of the capture radius was chosen to scale proportionally to the RMS end-to-end distance so as to keep the ratio R_c/R constant, the resulting dependences on the monomer positions (again, normalized by polymer length) for all polymers belonging to the same universality class collapsed to a single master curve (Fig. 4.4 demonstrates this for IE loops).

The precise shape of the curve depends on the universality class of the polymer chain. In particular, for short IE loops such that $j/N \ll 1$, the dependence of $T(0, j/N, R_c/R)$ on j/N is a power law, with a scaling exponent that is different in the case of Rouse chains (no excluded volume) and excluded-volume random coils (Fig. 4.4, inset). This extreme case is, however, inaccessible by our experimental measurements, where $0.3 < j/N \leq 1$. In the

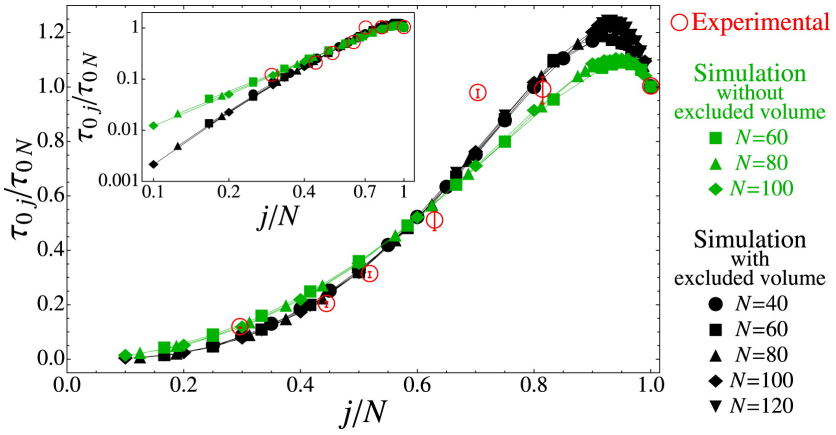


Figure 4.4: Universal dependence of the IE loop formation time on the monomer positions and the chain length. When the loop formation times are normalized by the end-to-end collision times and the ratio of the capture radius R_c to the end-to-end RMS distance, R, remains fixed (here, $R_c/R = 0.2$), the simulation data for chains of different lengths collapse onto a single curve. The precise shape of the curve, and, in particular, its power-law scaling with the loop length in the limit of short loops (*inset*), depends on the polymer's universality class (here we compare Rouse chain with and without excluded volume interactions). Remarkably, when normalized by the end-to-end loop formation time, the experimental data for loops in single-stranded DNA (same data as in Fig. 4.2) agrees quantitatively with the simulation results.

experimentally relevant regime, the excluded volume interactions have a relatively weak effect on the overall shape of the curve of $T(0, j/N, R_c/R)$ versus j/N. This finding is not significantly altered when hydrodynamic interactions within the polymer chain are taken into account: Although those interactions significantly alter the magnitudes of the un-normalized loop formation times, they only weakly affect the shape of the master curve (see the Supporting Material of Ref. [15]).

This weak effect of hydrodynamic interactions may seem somewhat sur-

prising, especially considering the existing theoretical view that the diffusion-controlled regime no longer holds after hydrodynamic effects are introduced [67, 70, 72, 97]. This apparent contradiction, however, is at least partly semantic: The operational definition of the diffusion-controlled limit adopted here is that the timescale of the process is proportional to the solvent viscosity. This definition is both appealing intuitively and testable experimentally [22, 37], but it is subtly different from that adopted in the literature [67, 70, 72, 97]. Specifically, those articles differentiate between the diffusion-controlled limit and what they refer to as the law-of-mass action (LMA) regime.

The latter regime, which is predicted to occur, e.g., for end-to-end collision kinetics when both excluded volume interactions and hydrodynamic effects are present, results in a loop formation rate that is proportional to the equilibrium probability of forming the loop. However, in our language, LMA regime is still diffusion-controlled as long as the proportionality factor exhibits inversely proportional solvent viscosity dependence, as in the Szabo-Schulten-Schulten theory [13, 22, 23].

A measurable signature of the LMA regime is a different value of the scaling exponent δ (compare to Eq. 4.1), although in practice this difference is relatively small and can be obscured by effects arising from the finite size of our polymers [70]. Indeed, the value of δ estimated from our simulations of excluded-volume chains in the presence of hydrodynamic interactions (see the Supporting Material of Ref. [15]) is in agreement with the LMA predictions [70, 97]. We thus conclude that our results agree with previous theoretical

views [67,70,72,97] and that the shape of the master curve observed in Fig. 4.4 is not significantly altered by the LMA regime, which still falls within our definition of diffusion-controlled kinetics.

Although the aforementioned dimensional analysis requires that the capture radius R_c be proportional to the polymer's RMS end-to-end distance in order to observe a chain length independent master curve, experimentally R_c is not a free parameter. Moreover, its precise value is generally unknown, thus introducing an uncertainty as to which value to use when comparing the theoretical master curve with experimental data. Fortunately, as anticipated above, the dependence of $T(i/N, j/N, R_c/R)$ on the capture radius is fairly weak.

Specifically, unlike power laws typically observed for the loop length dependence of the collision times, the dependence of $T(0, j/N, R_c/R)$ on R_c/R is logarithmic (Fig. 4.5). This weak, logarithmic dependence holds even for values of the capture radius that are unrealistically (unphysically) large; e.g., for values so large that the concept of a loop is no longer well defined. Of note, excluded volume effects lead to a stronger capture radius dependence as compared to that for Gaussian chains whereas hydrodynamic interactions do not have any significant effect (Fig. 4.5). We likewise note that in contrast to the rather weak dependence of the rescaled time, $T(0, j/N, R_c/R)$, the unnormalized loop formation time τ_{ij} exhibits a much stronger capture radius dependence (compare to Fig. S2 in the Supporting Material of Ref. [15]). This, again, highlights the advantage of rescaling the data according to Eq. 4.3

for a meaningful comparison between experiments and theory. We find that

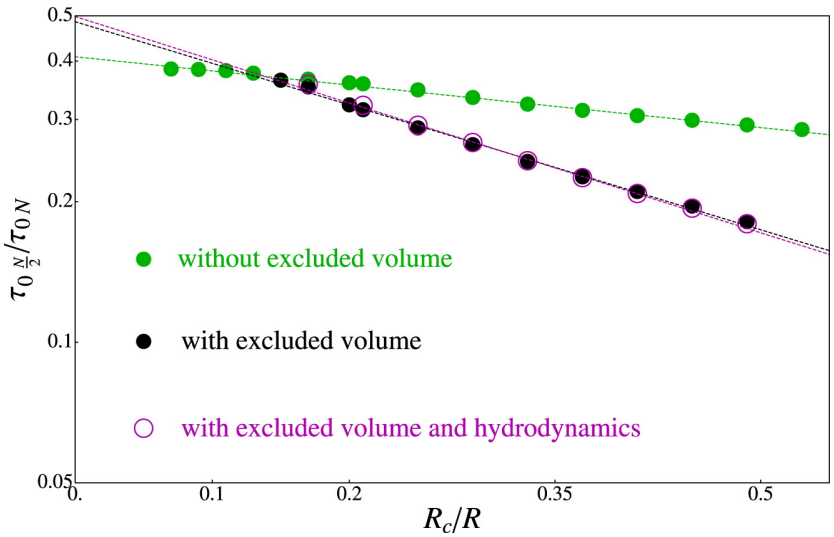


Figure 4.5: Effect of the capture radius on the internal-to-end loop formation kinetics: $T(i/N, j/N, R_c/R)$ has a relatively weak, logarithmic dependence on R_c/R. Here we illustrate this dependence for the loops formed between the middle of the chain and one of the chain ends, i.e., $i = 0$ and $j = N/2$. See the Supporting Material for Ref. [15] for further analysis of the capture radius effects.

the experimental data for single-stranded DNA also agrees with the above scaling conjecture. Indeed, normalized experimental loop formation times of ssDNA exhibits remarkably good agreement with the scaling predictions (Fig. 4.4). The only significant discrepancy between simulations and experiments is found when $j > 0.9N$; i.e., when the probe is placed within $\sim 10\%$ of end of the chain. That is, although the theoretically predicted dependence $T(0, j/N)$ shows a maximum near $j \sim (0.9 - 0.95) \times N$ that is similar to the rollover behavior that has been previously predicted for FRET-derived position-dependent chain reconfiguration times [106], the experimental data

plateaus for $j/N > 0.7$. This discrepancy likely originates from the limited flexibility of even single-stranded DNA, whose Kuhn segment is expected to consist of $n_K \sim 5 - 8$ monomers.

The last n_K nucleotides in our experimental chains thus constitute a nearly rigid segment that moves in a concerted fashion. We therefore expect τ_{0j} to be approximately the same for all j such that $j > N - n_K$. This effect is not captured by the simulations used to generate the dependences seen in Fig. 4.4, which employed a highly flexible chain of Kuhn length comparable with the monomer size. Thus, to experimentally observe the maximum occurring in our simulations at $j \sim (0.9 - 0.95) \times N$, it would be necessary to use DNA chains long enough that their tails (of length $0.05 - 0.1N$) are considerably longer than their Kuhn length. Limitations in the synthesis of the necessary DNA constructs, however, preclude experimental investigation of this hypothesis.

4.4.4 Intrachain dynamics in unfolded proteins: Comparison with simulations

The above-described comparison between theory and experiment requires knowledge of both the end-to-internal and end-to-end collision rates of each construct. In addition to our oligonucleotide studies, such data are also available for a limited set of polypeptide constructs. Specifically, Reiner et al. [100] have used triplet-triplet energy transfer to measure the diffusion-controlled rates of both internal and internal-to-end loops within an unfolded 36-residue long protein, the *villin headpiece*.

To our knowledge, their measurement of a loop between $i = 7$ and $j = 23$ is the only internal-to-internal collision rate reported in the literature to date. Unlike IE collisions, for which a plot of T versus j/N is sufficient for comparison between theory and experiment, a comparison with internal loops requires the entire surface of $\tau_{ij}(N)/\tau_{0N}(N)$ as a function of i/N and j/N. In doing so (Fig. 4.6) we find that, despite differences in the method used to probe the loop formation and in the physics of the polymer itself (unfolded protein versus single-stranded DNA), the data of Reiner et al. [100] also agree well with our predictions, with the exception of the $i = 23$, $j = 35$ pair, for which the corresponding loop formation time is considerably longer than expected. As pointed out by Reiner et al. [100], however, this anomaly presumably arises because of the residual helical order between these two residues. Indeed, a helix intervening between a pair of residues would preclude their short-ranged contact so that the measured loop formation time would be controlled by the timescale of helix unfolding [106, 108] and would be longer than for a fully random polymer.

4.5 Concluding Remarks

Life requires that its building blocks, biopolymers, populate highly specific, nonrandom conformational ensembles. Generic polymer chains, on the other hand, commonly exhibit universal scaling laws relating their various properties to their length. Those laws are independent of the details of intramolecular interactions, provided that the chains are long enough. Such

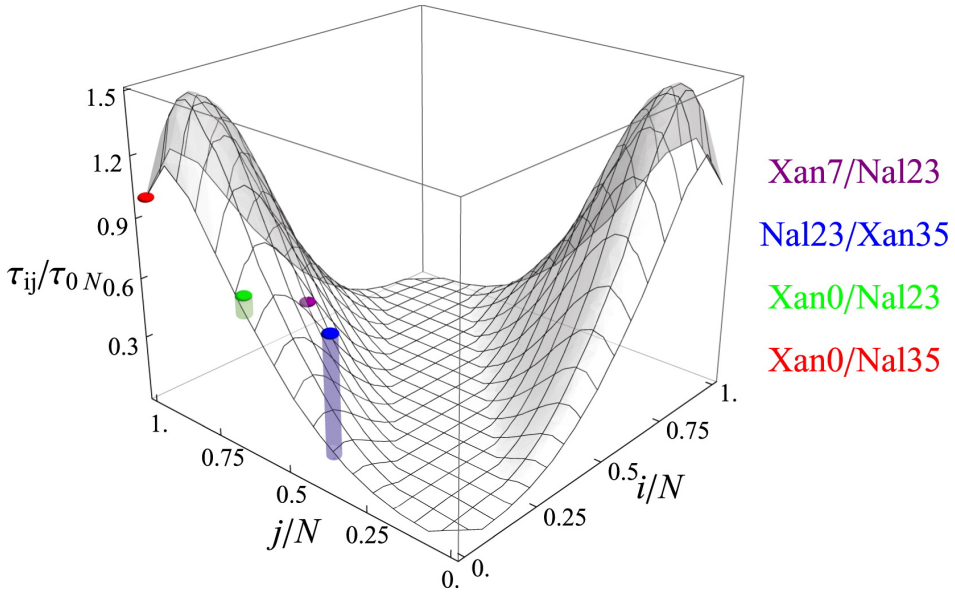


Figure 4.6: When rescaled according to Eq. 4.3, experimental measurements of internal loop formation times in an unfolded protein [100] agree with the simulation-derived dependence $T(i/N, j/N, R_c/R)$ (here we have assumed $R_c/R = 0.2$). The outlier observed for $i = 23$, $j = 35$, is due to residual helical order between those monomers.

polymers therefore provide a reference for measuring the structural organization within functional biomolecules. Unfortunately, comparison between universal scaling laws and the behavior of real biopolymers is obscured by finite-size effects: whereas scaling laws are only valid asymptotically in the infinitely long chain limit, biomolecules of interest are often not long enough to display such scaling laws or even to allow their reliable verification.

Here we have exposed a different type of general behavior in the dynamics of internal loop formation within disordered polymer chains and showed that the timescale of forming such loops, when properly renormalized, obeys

a simple universal dependence on the position of the loop extremities. In contrast to the chain length dependence, this dependence is in practice rather forgiving of the experimental constraints such as limited chain length or lack of microscopic knowledge of the probes used to make the measurements. At the same time, it is highly sensitive to any residual order found within the polymer chain, much more so than equilibrium properties such as the average distance between the monomers that form the loop. Indeed, average dimensions of unfolded proteins can exhibit Flory's random-coil scaling with chain length even when significant partial order is present [109, 110]. On the other hand, intervention of a rigid residual structure (e.g., a helix) between a pair of residues can suppress their respective collisions, thus considerably slowing down the observed loop formation time.

The robustness of the dependence of the loop formation time on the location of the loop-forming monomers established here for random coils on one hand, and its sensitivity to residual structural order on the other hand, suggest that spatio-temporal correlations inferred from measurements of timescales of interior loop formation in biopolymers can be used as a sensitive probe of both their structure and dynamics.

Chapter 5

Failure of one-dimensional Smoluchowski diffusion models to describe the duration of conformational rearrangements in floppy, diffusive molecular systems: A case study of polymer cyclization

5.1 Abstract

Motivated by recent experimental efforts to measure the duration of individual folding/unfolding transitions in proteins and RNA, here we use simulations to study the duration of a simple transition mimicking an elementary step in biopolymer folding: the closure of a loop in a long polymer chain. While the rate of such a transition is well approximated by a one-dimensional Smoluchowski model that views the end-to-end distance dynamics of a polymer chain as diffusion governed by the one-dimensional potential of mean force, the same model fails rather dramatically to describe the duration of such transitions. Instead, the latter timescale is well described by a model where the chain ends diffuse freely, uninuenced by the average entropic force imposed by the polymer chain. The effective diffusion coefcient then depends on the length scale of the loop closure transition. Our ndings suggest that simple one-dimensional models, when applied to estimate the duration of reactive events

in complex molecular systems, should be used with caution.

5.2 Introduction

A central quantity in traditional chemical kinetics is the rate k_{AB} at which a molecular transition from some subset of molecular conformations A (called the "reactants") to another subset B (called the "products") takes place. When the transitions between A and B are governed by first order kinetics, this rate is equal to the inverse of the mean first passage time for the system starting in A to arrive in B. If A and B are separated by a large free energy barrier, the transitions from A to B are rare and the associated timescale k_{AB}^{-1} is long, much longer than, say, typical timescales of molecular vibrations or other local rearrangements within A or B as it may take many failed attempts to cross the barrier until a successful one is encountered. In contrast, the duration of a successful transition event itself, which we refer to as the *transit time* t_{AB}, is typically much shorter than k_{AB}^{-1}. Measurement of transit times poses significant experimental difficulties and requires that molecular transitions be observed at a single-molecule level and with high time resolution. Nevertheless, recent improvements in the single-molecule data analysis have enabled several experimentalists to report transit time estimates for protein and RNA folding [111, 112]. Such measurements are particularly valuable because they provide a unique link between experimental observables and microscopic properties of transition paths, i.e., the trajectories followed by molecules as they undergo large-scale conformational changes. There has been great theoretical

interest in characterizing transition path ensembles in folding and in other conformational rearrangements occurring in complex molecular systems [113–120]. While an experimental counterpart for such studies is largely lacking, single-molecule transit time measurements perhaps offer one of the most promising efforts toward this goal.

The duration of molecular transitions has also received recent theoretical attention [121–126]. In particular, for simple barrier crossing problems it was shown [111, 121] that the mean transit time can be estimated as

$$\langle t_{AB} \rangle \sim \frac{1}{\omega_K} \log \left(\frac{\Delta V}{k_B T} \right) \tag{5.1}$$

where ω_K is the frequency of the unstable normal mode at the saddle point corresponding to the transition state and V is the typical energy barrier traversed by a transition path. The above result is likely applicable to reactions involving small molecules with the reactant, product, and transition state conformations corresponding to well defined structures. In contrast, the reactant state in, e.g., protein folding, involves a broad conformational ensemble of largely disordered polymer structures. Consequently, there is generally no single dominant pathway leading to the product state. More generally, pathways of large-scale conformational rearrangements in biomolecules often surmount significant entropic barriers and involve diffusive dynamics on rough or flat energy landscapes. Simple theoretical estimates such as the one provided by Eq. 5.1 and low-dimensional models employed in most studies [121–126] may not be applicable to such systems.

Here, we would like to understand what physical properties of a floppy, diffusive molecular system undergoing a large-scale conformational change determine the corresponding transit time. To this end, we study transit times in a model that captures many features of biomolecular transitions yet is not too taxing computationally: cyclization of a polymer chain (Fig. 5.1). As cyclization mimics the process by which distant parts of a polypeptide or polynucleotide chain come together to form its final structure, it also provides a simple example of "folding" [2–5, 33]. Moreover, just like more realistic folding phenomena, polymer cyclization lacks a single dominant pathway but rather exhibits a multitude of alternative paths leading to the same end result [127, 128].

Suppose we monitor the dynamics of the polymer chain and, specifically, the distance R between its ends, for a long time (Fig. 5.1). We define the reactant state A as all conformations with $R > R_A$. Starting in A, every once in a while the polymer will attain the product state B where the chain is in a cyclic form stabilized by a chemical bond (or other attractive interaction) between the chain ends. More precisely, this cyclic state is defined as the ensemble of all conformations with $R < R_B$. Our definitions of the reactant state A and the product state B mimic the typical experimental protocol, where the state of the system (e.g., whether or not the protein is folded) is inferred from some experimentally measured quantity (e.g., intensity of fluorescence emitted by a reporter molecule attached to the protein). For example, fluorescence resonance energy transfer measurements report the distance between two parts

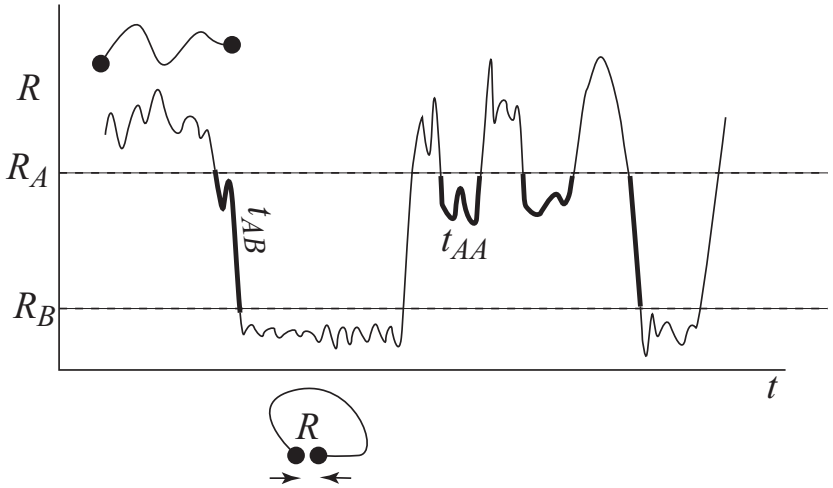


Figure 5.1: End-to-end distance trajectory of a polymer chain. Transition paths from A to B are the trajectory segments, where the end-to-end distance, R, enters the transition region at R_A and reaches R_B without first returning to R_A. The transit time t_{AB} is the time spent by a transition path within the transition region $R_B < R < R_A$. In contrast, trajectory segments that start at R_A and later exit the transition region into A prior to reaching R_B (duration t_{AA}) will contribute to the mean first passage time from A to B but not to the mean transit time.

of a molecular chain and can be used to differentiate between the folded and unfolded states of a protein [9, 111, 129, 130].

In a successful transition event, the system, starting at $R = R_A$, traverses the transition region, $R_B \leq R \leq R_A$, without ever recrossing back to the initial state A and ends up at $R = R_B$ (Fig. 5.1). The transit time t_{AB} is the duration of such an event. In the following sections, we will also compare this time with the first passage time, t_{AB}^{fp}, which is defined here as the time it takes the system to reach R_B provided that it has started at $R = R_A$. During t_{AB}^{fp}, the system is free to exit and re-enter the transition region (Fig. 5.1). There-

fore, this time may include the duration of many failed transition attempts and provides a measure of the cyclization rate.

Mean first passage times of polymer cyclization and, particularly, their dependence on chain length N have been the focus of numerous studies, see, e.g., Ref. [28] and references therein. A key finding reported here is that the transit timescales in such polymeric systems are fundamentally different from the first passage times. Specifically, while mean first passage times are essentially controlled by the global relaxation timescale of the chain [21], the mean transit time t_{AB} is governed by local chain rearrangements leading to an entirely different relationship between the mean transit time, the chain length, and the length scales defining the transition. Moreover, perhaps counterintuitively, typical transition paths in the cyclization process are consistent with a model where a small subset of n monomers ($n \ll N$) undergoes *free* diffusion unimpeded by the rest of the chain until the final configuration is attained.

Another message from this study concerns the use of low-dimensional models to describe dynamics of biomolecules. A common view is that those dynamics can be treated as one-dimensional motion in an effective potential of mean force, while the effect of the remaining degrees of freedom is captured by a phenomenological friction coefficient. Such models are widely used to describe various biomolecular phenomena, see, e.g., Refs. [14, 131–141]. In the context of polymer cyclization, this approximation is known as Szabo-Schulten-Schulten (SSS) theory [13, 23, 28]. Here, we however show that such an approximation, while reasonable for estimation of first passage times, fails,

even qualitatively, to predict transit times. Our results thus suggest that such one-dimensional theories, particularly when applied to the folding experiments [111], should be treated with suspicion. We note that limitations of one-dimensional models have also been pointed out in the context of mechanical protein unfolding experiments, see, e.g., Ref. [142].

5.3 Results

We have performed Langevin dynamics simulations of a bead-and-spring polymer model (see Section (5.5) for further details) and generated long time dependences of the polymer's end-to-end distance $R(t)$. The cyclic conformations are those for which the end-to-end distance is less than a specified value R_B. In our simulations, such cyclic conformations are short lived because no chemical bond was assumed to exist between the chain ends. A short ranged attractive potential acting between the ends at $R < R_B$ could mimic such a bond and would stabilize the cyclic form and thus increase the mean time the polymer spends being circular. However as long as this potential can be neglected for $R > R_B$, it will have no effect on the transit times from A to B, which are independent of the potential outside the transition region $R_B < R < R_A$ (see Ref. [121]). For this reason, no such potential was imposed in the simulations.

A transition event from A (i.e., the ensemble of conformations with end-to-end distances longer than R_A) to B (the ensemble of the cyclic conformations with $R < R_B$) starts whenever the end-to-end distance drops below

R_A and ends whenever it attains the value R_B without first returning to R_A. The transit time t_{AB} is the time such a transition lasts (Fig. 5.1). This time, of course, depends on the choice of both R_A and R_B. In practice, one hopes that this dependence would be weak enough that t_{AB} would report on an inherent property of the system under study rather than the specific way it is measured [111, 121]. Indeed, for simple models the distance dependence of the mean transit time is often logarithmic [111, 121]. Moreover, the long-time tail of the probability distribution of t_{AB} is independent of the spatial extent of the transition region [121].

The choice of the transition region boundaries is also restricted by physical considerations. A reasonable choice for R_A would be a value that is comparable with but is somewhat shorter than the polymer's mean end-to-end distance. Such a choice will ensure that the polymer conformations with $R > R_A$ are typical ones encountered in the equilibrium ensemble. Nevertheless, in order to better understand how transit times depend on the definition of the transition region, in the following R_A will be viewed as a free parameter not necessarily limited to physically sensible values.

5.3.1 Polymer cyclization transit times are independent of chain length

When the boundaries of the transition region, R_A and R_B are fixed, the mean transit time t_{AB} is independent of the number of chain monomers N (Fig. 5.2). This is in striking contrast with the strong chain length dependence

of the mean first passage time t_{AB}^{fp}, i.e., the mean time it takes to reach B having started from A (Fig. 5.2). Indeed, t_{AB}^{fp} scales as $N^2 - N^{2.2}$ with chain length (fits are not shown in Fig. 5.2), which is consistent with the $N^{1.5} - N^{2.2}$ scaling of the average polymer closure time predicted by various theories (see, e.g., Refs. [28] and [21]).

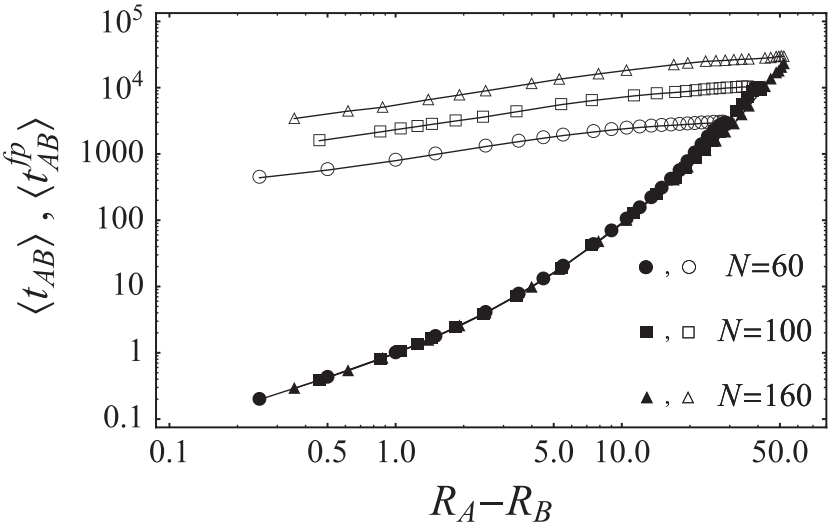


Figure 5.2: The mean transit time from A to B (filled symbols) is independent of the chain length N, in contrast to the strongly chain-length-dependent mean first passage from A to B (empty symbols). Here both times are plotted as a function of the spatial extent of the transition region $R_A - R_B$, with R_B fixed at 2.5 equilibrium bond lengths.

As a check of physical consistency, one also observes that for large values of R_A, the mean first passage times converge with the mean transit times (Fig. 5.2). This occurs because such large values of the end-to-end distance correspond to highly stretched polymer conformations. Because such conformations are rarely encountered in the course of polymer's dynamics, the

system originating in R_A is likely to reach R_B without ever returning to R_A and thus both the transit time and the first passage time are the same. Such large values are an example of an unphysical choice for R_A as they do not provide a physically meaningful boundary between typical noncyclic conformations and the cyclic ones.

5.3.2 1D Smoluchowski picture fails to describe the mean transit time

One of the common approaches to the problem of loop closure in polymer chains is the SSS theory [13], which assumes that the dynamics of the end-to-end distance R obeys a one-dimensional Smoluchowski diffusion equation in an effective potential of mean force given by $G(R) = -k_B T \ln p(R)$, where $p(R)$ is the equilibrium probability distribution of the end-to-end distance. Equivalently, this theory assumes that $R(t)$ obeys a one-dimensional overdamped Langevin equation without memory (see, e.g., Ref. [143]).

Although the deficiencies of this approximation, stemming from non-Markovian character of the end-to-end dynamics, have been recognized [12, 23, 26, 28], the SSS approximation nevertheless provides a useful tool that is often employed to interpret experimental data [22, 102]. Moreover, a generalization of this approximation recently proposed by Toan *et al.* [28] resolves at least some of its deficiencies.

In contrast, the same one-dimensional Smoluchowski model fails to even qualitatively reproduce the mean transit times observed here. Indeed, within

this model, the mean transit time is given by [125]

$$\langle t_{AB}\rangle = D^{-1}\int_{R_A}^{R_B}\exp(-G(R)/k_BT)\phi(R)(1-\phi(R))\,\mathrm{d}R\int_{R_A}^{R_B}\exp(G(R')/k_BT)\,\mathrm{d}R' \tag{5.2a}$$

$$\phi(R) = \int_{R}^{R_B}\exp(G(R')/k_BT)\,\mathrm{d}R'/\int_{R_A}^{R_B}\exp(G(R')/k_BT)\,\mathrm{d}R' \tag{5.2b}$$

where D is the effective diffusion constant. A simple estimate for D is obtained by matching the autocorrelation function for the end-to-end distance R estimated from the one-dimensional diffusion model and that from the full-blown dynamics of the original polymer chain [15, 74]. Assuming Rouse model for polymer dynamics [44] (i.e., Gaussian chain statistics and no hydrodynamic interactions), this leads to a diffusion coefficient that is inversely proportional to the number of monomers in the chain,

$$D = \frac{\pi^2 k_B T}{4\gamma_0 N} \tag{5.3}$$

A comparison of the simulated transit times with Eq. 5.2 (Fig. 5.3) shows the failure of the 1D Smoluchowski model to describe transit times for polymer cyclization. In particular, Eq. 5.2 predicts t_{AB} to be chain length dependent, consistent with the fact that both the potential of mean force $G(R)$ and the diffusion coefficient depend on N. Moreover, assuming a different chain length dependence of the diffusion constant does not reconcile the Smoluchowski model with our data. For example, a chain-length independent value of D postulated by some theories [23] still leads to chain-length dependent transit times (data not shown) because the potential of mean force still depends on N. Even if one assumes that the chain length dependence of the

diffusion coefficient, $D(N)$, can be such that it would miraculously cancel the effect of the chain length on the potential of mean force $G(R)$, the dependence of t_{AB} on the spatial extent of the transition region, $R_A - R_B$, as predicted by the Smoluchowski model will still disagree with our data. Indeed, consider the logarithmic plot of t_{AB} as a function of $R_A - R_B$, as in Fig. 5.3. If, for some particular value of the chain length N, one uses a different value of the diffusion coefficient D, this will simply shift the curve predicted by Eq. 5.2 in the vertical direction, since D enters into this equation as a constant factor. It is obvious from Fig. 5.3 that no such shift can reconcile Eq. 5.2 with our data. This observation suggests that the problem with using the Smoluchowski theory lies in the assumption that transition paths are effectively subjected to the deterministic potential $G(R)$.

5.3.3 Free-diffusion model for the transition paths in polymer cyclization

While the model where the chain ends undergo one-dimensional diffusion biased by the potential of mean force $G(R)$ disagrees with the transit times observed for polymer cyclization, our data is quantitatively described by a model in which the chain extremities undergo free diffusion. To explain the physical basis for this model, consider a typical configuration of the chain that has just entered the transition region so that $R = R_A$ (Fig. 5.4). A global reconfiguration of the entire chain is not required to attain a cyclic chain conformation. Instead, closure of the loop can be accomplished by moving

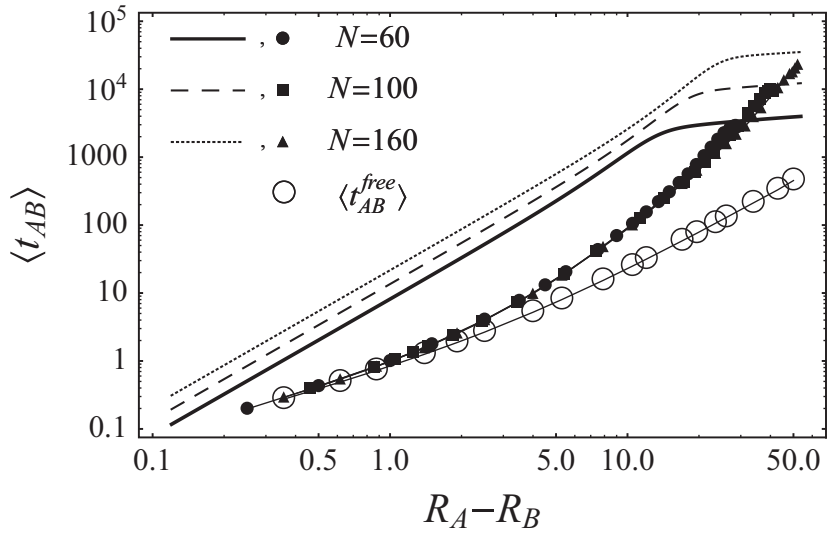


Figure 5.3: The mean transit times for cyclization of a polymer chain (filled circles) are independent of the chain length. For small values of $R_A - R_B$ they agree with the free diffusion model (open circles), which ignores the polymer chain and assumes that the end monomers move freely. In contrast, the model of Eq. 5.2 (solid, dashed, and dotted lines) that assumes 1D diffusion in a one-dimensional potential of mean force is inconsistent with the polymer data and shows significant dependence on the chain length. As in Fig. 5.2, the distance R_B is fixed here at 2.5 equilibrium bond lengths.

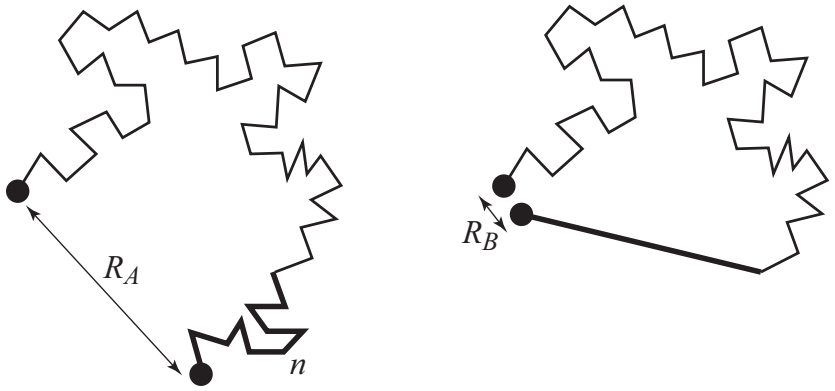


Figure 5.4: Cyclization of a polymer chain can be achieved via straightening of a chain segment containing $n \ll N$ monomers.

roughly

$$n(R_A, R_B) \sim (R_A - R_B)/\sigma \tag{5.4}$$

monomers, where σ is the bond length between successive monomers. Indeed, straightening the outer segment of the chain containing n monomers will result in a change in the end-to-end distance of order $\sigma n - \sigma\sqrt{n} \approx \sigma n$, which is equated to $R_A - R_B$ to produce Eq. 5.4. If roughly $n/2$ monomers move at each end of the chain, the effective frictional drag coefficient on each moving entity is proportional to $\gamma_0 n/2$, where γ_0 is the monomer friction coefficient. We thus envision two freely diffusing super-beads, each with an effective friction coefficient,

$$\gamma_{eff}(R_A, R_B) = \gamma_0 n(R_A, R_B)/2 \tag{5.5}$$

and consider transition paths in which the distance between the two changes from R_A to R_B. The mean transit time and its distribution is straightforward to compute either by solving a three-dimensional diffusion equation

with absorbing boundary conditions at the boundaries of the transition region [121,126] or via overdamped Langevin dynamics simulations for two noninteracting beads (the latter option has been chosen here). Because the times of Brownian motion scale directly proportional to the friction coefficient, the free diffusion model implies that the mean transit time should be given by

$$\langle t_{AB} \rangle = \frac{\gamma_{eff}(R_A, R_B)}{\gamma_0} \langle t_{AB}^{free}(R_A, R_B) \rangle = \frac{n(R_A, R_B)}{2} \langle t_{AB}^{free}(R_A, R_B) \rangle \qquad (5.6)$$

where t_{AB}^{free} is the transit time computed for two noninteracting monomers. It further implies that the probability distribution of the transit time, $p(t_{AB})$, can be obtained by simply rescaling the distribution $p_{free}(t_{AB}^{free})$ for a pair of noninteracting monomers,

$$p(t_{AB}) = (\gamma_0/\gamma_{eff}) p_{free}(t_{AB}\gamma_0/\gamma_{eff}) \qquad (5.7)$$

The dependence of the mean transit time on the distance $R_A - R_B$ compared to that estimated for a pair of freely diffusing monomers shows that for sufficiently short distances the two times are identical (Fig. 5.3), suggesting that the typical transition paths are local motions of the end monomers not involving the rest of the chain. As the distance $R_A - R_B$ is increased, more monomers have to be involved [cf. Eq. 5.4] in a typical transition and so the mean transit time becomes longer than that for unconstrained end monomers.

Instead of using the rough estimate of Eq. 5.4, let us now view Eqs. 5.5 and 5.6 as the definition of the mean number of monomers $n = n(R_A, R_B)$ engaged in a transition path. Then we can calculate its dependence on the

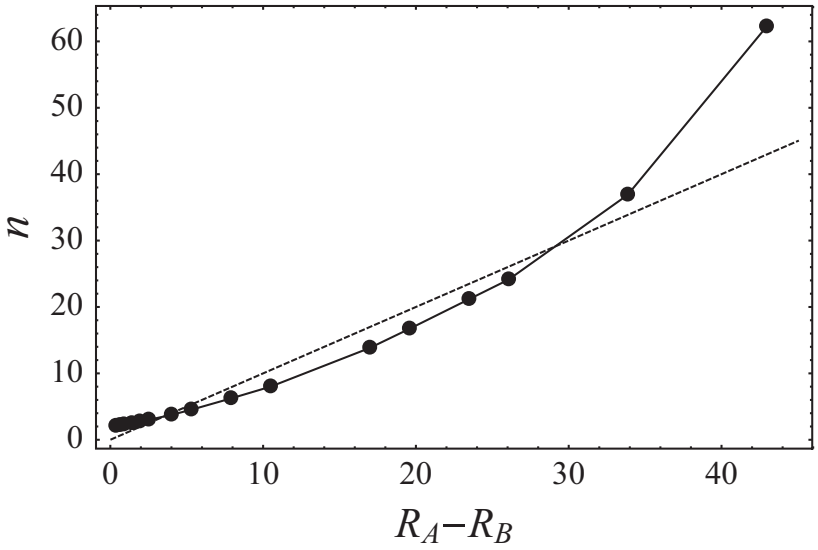


Figure 5.5: The effective number of monomers rearranging in a cyclization transition estimated from Eqs. 5.5 and 5.6. The effective number of monomers $n(R_A, R_B)$ involved in the cyclization of a chain increases with the change in the end-to-end distance $R_A - R_B$ required to close the loop. For small values of $R_A - R_B$ loop closure is accomplished by moving only $n(R_A, R_B) = 2$ end monomers. Here R_B is fixed at 2.5 bond lengths. Dashed line shows Eq. 5.4, which predicts that $n(R_A, R_B)$ is proportional to the distance $R_A - R_B$ traveled during a transition.

boundaries of the transition region, R_A and R_B, from Eq. 5.6. The result is shown in Fig. 5.5. At short distances, only the end monomers have to move so that we have $n(R_A, R_B) = 2$. For longer distances, we have $n(R_A, R_B) > 2$. Although in contrast to the naive estimate of Eq. 5.4, the dependence of $n(R_A, R_B)$ on $R_A - R_B$ is not exactly linear, Eq. 5.4 still predicts the effective friction coefficient reasonably well, except for very large values of $R_A - R_B$ (Fig. 5.5).

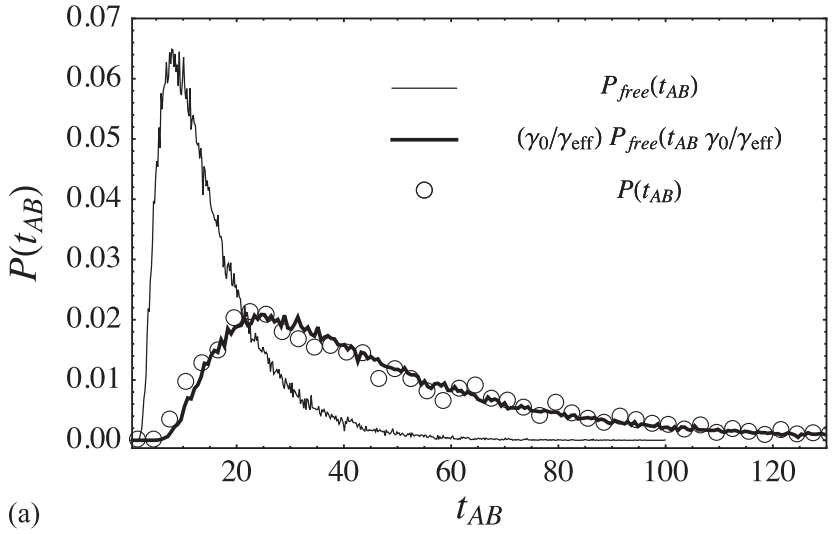
Further support of the free diffusion model comes from its accurate

prediction for the distribution of the transit times, $p(t_{AB})$. Indeed, using the effective number of monomers $n(R_A, R_B)$ as the only adjustable parameter and rescaling $p_{free}(t_{AB}^{free})$ according to Eqs. 5.5 and 5.7, we find the free diffusion model to predict, nearly perfectly, the transit time distributions $p(t_{AB})$ for different values of R_A and R_B (Fig. 5.6).

5.4 Concluding Remarks

Low dimensional models are widely accepted to describe complex phenomena in condensed phase chemical kinetics. In such models, a small subset of relevant degrees of freedom is identified, while others are presumed to produce an effective potential of mean force. The dynamic effect of the degrees of freedom that are left out of the picture is to produce randomly fluctuating forces and a frictional force, which is often described in terms of phenomenological friction coefficients. An archetypal example of this kind is the famous Kramers model introduced some 70 years ago [141]. This model considers thermally activated transitions over a one-dimensional barrier in a noisy, dissipative system and remains to be a standard description of barrier crossing phenomena in condensed-phase systems. Since most experimental data is inherently one- or low-dimensional, this model is especially valuable as a means of interpreting experiments.

The results reported here, however, highlight the limitations of simple low-dimensional models for a description of complex transitions. Indeed, we have found that the end-to-end dynamics of a polymer, as revealed by typical



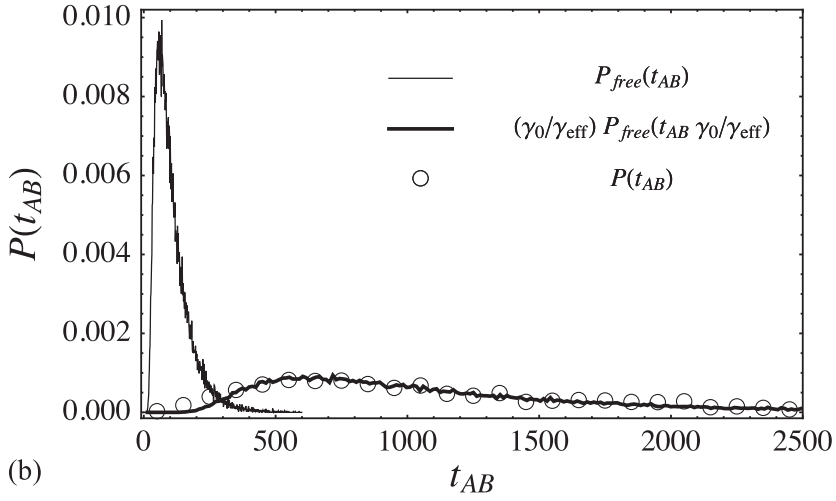


Figure 5.6: The distribution of transit times obtained from the free diffusion model, when rescaled according to Eq. 5.7 to account for the effective number of monomers involved, is nearly identical to the actual distribution obtained from simulations of polymer chains. (a): $N = 160$, $R_A = 10.39$, and $R_B = 2.5$. (b): $N = 160$, $R_A = 25.97$, and $R_B = 2.5$.

loop closure transition paths, is uninfluenced by the average entropic forces acting between the two ends of the polymer chain. This finding is not entirely surprising. The ensemble of transition paths consists only of those trajectories that have successfully made it directly from R_A to R_B (Fig. 5.1) and is fundamentally different from the ensemble of all possible paths. A similar observation has been made in our earlier study of a toy model involving transitions in a two-dimensional potential [121], which showed that the mean transit time is determined by a typical energy barrier V traversed by transition paths [Eq. 5.1]. This barrier is generally different from the free energy barrier surmounted in the transition. The difference between the two types of barriers is more dramatic in the present study because the free energy barrier encountered in cyclization of a flexible polymer model is almost entirely entropic, while the underlying energy landscape is flat. The free diffusion model has the same flat energy landscape and so it successfully describes the distribution of the transition path sub-ensemble. The same free diffusion model would be completely wrong if one were to attempt using it to calculate the mean first passage time to go from R_A to R_B, because it would not properly account for the entropic cost to initiate a successful transition path.

Because entropic contributions to the free energy barrier for protein folding are significant (see, e.g., Ref. [128] and references therein), it is then likely that Kramers-type models would be inadequate for a quantitative description for the duration of protein folding/unfolding events. This issue deserves further exploration with more realistic protein models.

Another interesting feature that emerges from our study is scale-dependent effective friction [Eqs. 5.5, 5.6, 5.7]. The number n of monomers involved in a transition path depends on the length scale of the polymer rearrangement, resulting in an effective friction whose value depends on the length of the path that needs to be traveled to close the loop. Such scale-dependent effective friction stems from non-Markovian end-to-end dynamics. At short time- or length-scales this dynamics looks like free monomer diffusion, while at long timescales it involves global rearrangement of the entire chain [20]. A similar length-scale dependence has been discussed by Toan *et al* [28] in their study of the mean loop closure time. In fact, if we assume that the typical distance at the beginning of a transition, R_A, scales as $N^{1/2}$, and that $R_A \gg R_B$, then our Eq. 5.4 predicts the effective number of monomer engaged in a typical transition to be proportional to $N^{1/2}$. This leads to an effective end-to-end diffusion constant that is inversely proportional to $N^{1/2}$, in agreement with the generalized SSS theory of Toan *et al.* proposed for the case of Rouse polymer chains [28]. The arguments that have led us to introduce the effective friction coefficient in the preceding section thus offer an intuitive explanation of the chain-length dependence of the effective friction proposed in Ref. [28]: It is simply proportional to the minimum number of monomers that need to be displaced in order for the chain ends to meet. We emphasize, however, that while the effective friction coefficient found here is consistent with that predicted in Ref. [28], the effective potential appropriate for the description of the polymer's end-to-end motion in a typical transition event is a different

one, being flat and unaffected by the entropic forces between the chain ends.

5.5 Appendix: Simulation details

Our polymer model consists of $N+1$ beads of mass m connected by N springs. Consecutive (bonded) beads separated by a distance r interact via a harmonic potential of the form

$$V_{bond}(r) = k_{bond}(r-\sigma)^2/2 \tag{5.8}$$

where σ is the equilibrium bond length and $k_{bonded} = 100.0\epsilon/\sigma^2$. Excluded volume interactions between nonconsecutive (non-bonded) beads separated by a distance r are represented by a repulsive Lennard-Jones potential,

$$V_{non-bonded}(r) = 4\epsilon[(\sigma/r)^{12} - (\sigma/r)^6]\theta(2^{1/6}\sigma - r) \tag{5.9}$$

where ϵ is a characteristic energy scale and θ is the Heaviside step function that truncates the attractive portion of the Lennard-Jones potential. The dynamics of each bead is governed by the Langevin equation

$$m\frac{d^2\mathbf{r}_i}{dt^2} = -\frac{\partial V}{\partial \mathbf{r}_i} - \gamma_0\frac{d\mathbf{r}_i}{dt} + \Omega_i(t) \tag{5.10}$$

where $\mathbf{r}_i(t)$ is the position of the bead, V is the total interaction potential, γ_0 is the friction coefficient of a bead (monomer), and Ω_i is a random force satisfying the fluctuation-dissipation theorem. A friction coefficient of $\gamma_0 = 2.0(\sigma^2/m\epsilon)^{-1/2}$ was used to ensure that the dynamics are in the overdamped regime and simulations were performed at a temperature of $T = 1.0\epsilon/k_B$. Lengths and times are reported in this study in dimensionless units of σ and $\tau_0 = (m\sigma^2/\epsilon)^{1/2}$, respectively.

Chapter 6

Exploring the role of internal friction in the dynamics of unfolded proteins using simple polymer models

6.1 Abstract

Using simple polymer models, we explore the role of internal friction in two common single molecule experiments of unfolded proteins: end-to-end reconfiguration dynamics and end-to-end loop formation. We use the Rouse with internal friction (RIF) model as the framework for our studies due to its incorporation of internal friction as an adjustable free parameter, allowing for a systematic study of internal friction, as well as its applicability to unfolded state dynamics which was recently demonstrated in a related study [Soranno et al. PNAS 2012 [144]]. Furthermore, we explore the conformational dynamics under non-free draining conditions using a new model: Zimm with internal friction (ZIF). We construct several experimentally testable predictions using our simple models. In the context of reconfiguration dynamics, we find that internal friction provides an additive contribution to the reconfiguration times, a result that was previously discussed for the RIF model. However, we find that the picture is more complicated in loop formation times, where internal friction is not simply additive nor properly described by any simple relation. Instead,

we perform a detailed study of the loop formation time in the asymptotic limits of very high and very low internal friction, finding that an exact solution can be written for the former and an approximate scaling relation can be used to describe the latter.

6.2 Introduction

The timescales at which unfolded and nonnative conformations of proteins diffusively undergo re-arrangement are intimately related to the "speed limit" of protein folding [1] and, thus, have consequently attracted a great deal of attention from the scientific community. To this end, the simple polymer model of Rouse [43, 44] has been used to help characterize the reconfiguration dynamics of unfolded proteins that exhibit random-coil dimensions [106]. An even simpler description is offered by Kramers-like theories [13, 31], where, for example, the rate of collisions between the ends of a polymer is viewed as diffusion over a barrier along a single reaction coordinate. In using simple polymer models or Kramers-like theories, it is commonly assumed that solvent-mediated friction is the only source of frictional drag on the molecule of interest. This assumption is, however, flawed since not all amino acids may have the same degree of solvent exposure, and thus, other dissipative mechanisms (i.e., internal friction) may play a significant role in their dynamics.

The concept of internal friction reflects the resistance of a polymer to changes in its conformation and is often interpreted as reflecting the roughness of the underlying energy landscape [145]. Possible mechanisms that contribute

to internal friction include dihedral angle rotational barriers [146, 147], hydrogen bonding [148] and intrachain collisions [149], suggesting that internal friction would perhaps be more pronounced in the native state or other compact conformations where, for example, hydrogen bonding and steric barriers are more prevalent. However, the importance of internal friction in the unfolded state dynamics and its effect on the protein folding speed limit remains an open issue. Previous experimental studies have attempted to quantify internal friction in the native state [150] and in the protein folding kinetics [149, 151–154] using phenomenological models to describe this effect. A recent single-molecule FRET study of the reconfiguration dynamics of unfolded *T. maritime* cold shock protein and intrinsically disordered proteins [144] provided a more quantitative account of internal friction in terms of a well established model of internal friction called Rouse with internal friction (RIF) [43, 155]. This study suggests that simple polymer models such as RIF may provide sufficient framework to describe the dynamics of the unfolded state. Of course, extensions of those studies to other proteins and other types of experimental measurements would be needed to further validate this view. In anticipation of such studies, this paper further explores experimentally verifiable predictions of RIF and related models.

The RIF model is not new and has also been used to study strongly stretched biopolymers [155, 156] and to demonstrate the complex interplay between the solvent friction and internal friction in the dynamics of polypeptides [157]. Given its simplicity that often affords analytic solutions, it provides

an attractive alternative to the computationally costly molecular dynamics simulations, although the latter would still be required to elucidate molecular mechanisms of internal friction. In addition to RIF, here we also consider a new model that is the non-free draining version of RIF: Zimm with internal friction (ZIF). Like RIF, ZIF employs a single parameter to quantify the magnitude of internal friction and retains the simplicity of the mathematical structure of the RIF solution. However, unlike the free-draining Rouse models, the more realistic ZIF builds upon the Zimm model [44], which accounts for hydrodynamic interactions among the chain monomers.

Using RIF, ZIF and their extensions, we attempt to mimic typical experimental studies that probe the unfolded state dynamics with the goal of establishing how internal friction manifests itself in these types of experiments. One such experiment uses single-molecule nanosecond fluorescence correlation spectroscopy (nsFCS), which employs fluorescence resonance energy transfer (FRET) between a pair of dyes attached at different locations along the backbone of a peptide. Fluctuations in the fluorescence intensities of the donor and acceptor, as related to the fluctuations in the distance between the FRET pairs, are monitored and a characteristic timescale for these fluctuations is obtained−i.e., a reconfiguration time [9–11]. Alternatively, a fluorescent probe and a quencher can be attached at different positions on a peptide chain. In this case, the time it takes for the probe and quencher to diffusively come into contact can be obtained−i.e., loop formation time [2,4,35,37]. The timescales inferred from the two types of experiments are not the same, although they

are related to each other. Indeed, fluorescence quenching experiments probe the time to the formation of a close contact between the quencher and the probe and, consequently, yield longer timescales than those exhibited in fluctuations of long-range energy transfer employed in nsFCS measurements. It has been shown that by varying the relative positions of the FRET pairs or the probe/quencher pair in the two respective experiments, one can probe the spectrum of relaxation times of a polymer chain [11, 15, 106]. Here, however, for simplicity we mainly focus on the role of internal friction in the end-to-end dynamics.

This paper is organized as follows. Section 6.3 will discuss the models used in our study as well as the simulation details. Section 6.4 will discuss the role of internal friction in the reconfiguration dynamics of our simple models. Section 6.5 will discuss the role of internal friction in the loop formation dynamics of our simple models and provide a comparison with the experimental loop formation times of a peptide. Finally, in Section 6.6 we will conclude with a discussion of our primary findings.

6.3 Model details

Consider a linear polymer chain that consists of $N+1$ monomers with coordinates $\mathbf{r} = (\mathbf{r}_0, \mathbf{r}_1, ..., \mathbf{r}_N)$. To describe the dynamics of this chain, we start with the equation of motion for a polymer composed of Brownian particles in

an incompressible fluid with low Reynolds number [44]:

$$\frac{d\mathbf{r}}{dt} = \mathbf{HF} \tag{6.1}$$

where $\mathbf{F} = (\mathbf{F}_0, \mathbf{F}_1, ..., \mathbf{F}_N)$ represents the sum of forces on each monomer and $\mathbf{H}$ is the mobility tensor with components $\mathbf{H}_{nm} \equiv \mathbf{H}(\mathbf{r}_n - \mathbf{r}_m)$ that describes the hydrodynamic interaction (Note: $\mathbf{H}_{nm}$ is a 3 by 3 matrix and not a scalar component).

6.3.1 Rouse and Rouse with internal friction (RIF)

The Rouse model [43, 44] describes a linear polymer where neighboring monomers along the chain are connected by harmonic springs and the excluded volume interaction is neglected. The hydrodynamic interaction is neglected such that $\mathbf{H}_{ij} = \mathbf{I}\delta_{ij}/\xi_s$, where ξ_s is the solvent friction coefficient. Consequently, Eq. 6.1 for the Rouse model is expressed as

$$\xi_s \frac{d\mathbf{r}}{dt} = \kappa \mathbf{k}\mathbf{r} + \mathbf{f} \tag{6.2}$$

where $\kappa = 3k_BT/b^2$ is the stiffness of the connecting springs, b is the Kuhn length, $\mathbf{f} = (\mathbf{f}_0, \mathbf{f}_1, ..., \mathbf{f}_N)$ are the random forces on each monomer and $\mathbf{k}$ is the tri-diagonal connectivity matrix:

$$\mathbf{k} = \begin{pmatrix} -1 & 1 & 0 & 0 & \cdots \\ 1 & -2 & 1 & 0 & \cdots \\ 0 & 1 & -2 & 1 & \cdots \\ 0 & 0 & 1 & -2 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \tag{6.3}$$

The phenomenological incorporation of internal friction in the RIF model [43, 155] is accomplished by introducing a frictional drag term to the Rouse model

that resists the relative extension between connected monomers. The resultant equation of motion for the RIF model is

$$\xi_s \frac{d\mathbf{r}}{dt} = \kappa \mathbf{k}\mathbf{r} + \xi_i \frac{d}{dt}\mathbf{k}\mathbf{r} + \mathbf{f}^* \tag{6.4}$$

where ξ_i is the internal friction coefficient and $\mathbf{f}^*$ is the appropriate random force vector [155].

Similarly to Eq. 6.2 (Rouse), Eq. 6.4 (RIF) can be simplified by making a coordinate transformation into the basis of the eigenvectors of the connectivity matrix−i.e., normal mode decomposition. The eigenvectors of $\mathbf{k}$ are given by

$$u_{pn} = \frac{1}{N+1} \cos\left(\frac{p\pi}{N+1}\left(n + \frac{1}{2}\right)\right) \tag{6.5}$$

and have the associated eigenvalues

$$\lambda_p = 4\sin^2\left(\frac{p\pi}{2(N+1)}\right) \tag{6.6}$$

where $p \in [0, N]$ denotes the mode index and $n \in [0, N]$ once again denotes the monomer index. The dynamics of the polymer chain can thus be written as a superposition of independent eigenmodes:

$$\mathbf{r}(t) = \sum_{p=0}^{N} \mathbf{X}_p(t)\mathbf{u}_p \tag{6.7}$$

Transformation of Eq. 6.4 with the eigenvectors of Eq. 6.5 results in $N+1$ independent overdamped harmonic oscillator equations of motion:

$$\xi_p^{(RIF)} \frac{d\mathbf{X}_p}{dt} = -k_p \mathbf{X}_p + (2\xi_p^{(RIF)} k_B T)^{1/2} \mathbf{a}_p(t) \tag{6.8}$$

For the p-th mode, $\xi_p^{(RIF)} = \xi_s + \xi_i\lambda_p$ is the friction coefficient, $k_p = \kappa\lambda_p$ is the spring coefficient (identical to that of the Rouse model), and $\mathbf{a}_p = (a_{p,x}, a_{p,y}, a_{p,z})$ is a three-dimensional Gaussian process characterized by the moments $\langle \mathbf{a}_p(t) \rangle = 0$ and $\langle a_{p\alpha}(t) a_{q\beta}(t') \rangle = \delta_{pq}\delta_{\alpha\beta}\delta(t-t')$ where p and q denote the mode index while α and β denote the x, y, or z directions. Furthermore, the relaxation time of each RIF mode is written as

$$\tau_p^{(RIF)} = \frac{\xi_p^{(RIF)}}{k_p} \cong \frac{\tau_{Rouse}}{p^2} + \tau_i \ , \ p > 0 \tag{6.9}$$

where $\tau_{Rouse} = \xi_s N^2 b^2/(3\pi^2 k_B T)$ is the slowest relaxation time of the Rouse model and $\tau_i = \xi_i/\kappa$ is the relaxation time due to internal friction. In the absence of internal friction (i.e., $\xi_i = 0$), the equation of motion for each independent mode and the spectrum of relaxation times of the Rouse model are recovered from Eq. 6.8 and Eq. 6.9, respectively.

6.3.2 Zimm with internal friction (ZIF)

The Zimm model [44] accounts for the hydrodynamic interaction using the pre-averaging approximation where $\mathbf{H}_{nm}$ is replaced by its average:

$$\mathbf{H}_{nm} \rightarrow \langle \mathbf{H}_{nm} \rangle = \frac{\mathbf{I}}{2\pi^2 \xi_s |n-m|^{1/2}} \tag{6.10}$$

The friction coefficient is related to the solvent viscosity by $\xi_s = 3\pi\eta_s b$ (i.e., each monomer is treated as a sphere with a hydrodynamic radius of $b/2$ undergoing Stokes flow). To construct the equation of motion for the Zimm with internal friction (ZIF) model, we combine Eq. 6.10 and Eq. 6.4 into Eq. 6.1:

$$\frac{d\mathbf{r}}{dt} = \mathbf{H}\left(\kappa \mathbf{k}\mathbf{r} + \xi_i \frac{d}{dt}\mathbf{k}\mathbf{r} + \mathbf{f}^*\right) \tag{6.11}$$

Similarly, Eq. 6.11 can be simplified by making a coordinate transformation into the basis of the eigenvectors of matrix $\mathbf{k}$ given by Eq. 6.5:

$$\xi_p^{(ZIF)} \frac{d\mathbf{X}_p}{dt} = -k_p \mathbf{X}_p + (2\xi_p^{(ZIF)} k_B T)^{1/2} \mathbf{a}_p(t) \tag{6.12a}$$

$$\xi_p^{(ZIF)} = \xi_s \left(\frac{\pi p}{3(N+1)} \right)^{1/2} + \xi_i \lambda_p \ , \ p > 0 \tag{6.12b}$$

$$\xi_0^{(ZIF)} = \xi_0^{(Zimm)} = \xi_s \left(\frac{3\pi}{32N} \right)^{1/2} \ , \ p = 0 \tag{6.12c}$$

For the p-th mode, $\xi_p^{(ZIF)}$ is the friction coefficient, $k_p = \kappa \lambda_p$ is the spring constant, and $\mathbf{a}_p = (a_{p,x}, a_{p,y}, a_{p,z})$ is, again, a three-dimensional Gaussian process characterized by the moments $\langle \mathbf{a}_p(t) \rangle = 0$ and $\langle a_{p\alpha}(t) a_{q\beta}(t') \rangle = \delta_{pq} \delta_{\alpha\beta} \delta(t - t')$ as in Eq. 6.8. Likewise, the relaxation time of each ZIF mode is written as

$$\tau_p^{(ZIF)} = \frac{\xi_p^{(ZIF)}}{k_p} \cong \frac{\tau_{Zimm}}{p^{3/2}} + \tau_i \ , \ p > 0 \tag{6.13}$$

where $\tau_{Zimm} = \xi_s (N+1)^{3/2} b^2 / (3\sqrt{3} k_B T \pi^{3/2})$ is the slowest relaxation time of the Zimm chain and $\tau_i = \xi_i / \kappa$ is the relaxation time due to internal friction. In the absence of internal friction (i.e., $\xi_i = 0$), the equation of motion for each independent mode and the spectrum of relaxation times of the Zimm model are recovered from Eq. 6.12a and Eq. 6.13, respectively. It is well known that the Zimm model is able to correctly reproduce the chain length dependence of both the translational diffusion coefficient (i.e., $D_{chain} = k_B T / (N \xi_0^{(Zimm)}) \sim N^{-1/2}$) and the relaxation times ($\tau_{Zimm} \sim N^{3/2}$) of a polymer under theta conditions [44].

6.3.3 Simulation details

Simulation of our simple models were performed by carrying out Brownian dynamics integration of the $N+1$ three-dimensional overdamped harmonic oscillator equations of motion represented by Eq. 6.8 and Eq. 6.12a. The positions of the monomer ends as a function of time were then recovered through the linear transformation of Eq. 6.7 to obtain a trajectory of end-to-end distance, $R(t)$. The loop formation time, τ_{loop}, was computed from the trajectory $R(t)$ as a mean first passage time for the ends of a polymer chain to diffusively come within a distance R_c of one another (i.e., $R < R_c$) when starting from the equilibrium ensemble of conformations with $R > R_c$ [15]. In our study, the unit of length is the Kuhn length b, the unit of time is $\tau_0 = b^2/D_0$ where D_0 is the diffusion coefficient of a monomer, and the unit of energy is k_BT.

6.4 Reconfiguration times of simple models

6.4.1 Definition of reconfiguration time

Consider the autocorrelation function of the distance vector between a pair of monomers i and j:

$$\phi_{ij}(t) = \langle (\mathbf{r}_i(t) - \mathbf{r}_j(t)) \bullet (\mathbf{r}_i(0) - \mathbf{r}_j(0)) \rangle \tag{6.14}$$

A reconfiguration time, τ_{ij}, can be defined from the relaxation of $\phi_{ij}(t)$ [106, 144]:

$$\tau_{ij} = \int_0^\infty \mathrm{d}t \phi_{ij}(t)/\phi_{ij}(0) \tag{6.15}$$

This definition of a reconfiguration time is unique if $\phi_{ij}(t)$ exhibits exponential decay—i.e., $\phi_{ij}(t) = \phi_{ij}(0)\exp(-t/\tau_{ij})$. However, a reconfiguration time can still be defined as Eq. 6.15 if $\phi_{ij}(t)$ is non-exponential, although such a definition would no longer be unique [158].

A reconfiguration time defined from Eqs. 6.14 and 6.15 can be evaluated analytically from our simple models by substituting in the linear combinations of the eigenmodes from Eq. 6.7:

$$\tau_{ij} = \frac{\xi_0}{\kappa}\left(\frac{\sum_{p=1}^{N}(\lambda_p)^{-2}(u_{pi}-u_{pj})^2}{\sum_{p=1}^{N}(\lambda_p)^{-1}(u_{pi}-u_{pj})^2}\right) + \tau_i \tag{6.16}$$

where ξ_0 is the friction coefficient of the translational mode which is equal to $\xi_0^{(Rouse)} = \xi_s$ for the RIF model and $\xi_0^{(Zimm)} \propto \xi_s$ (given by Eq. 6.12a) for the ZIF model. We will refer to the reconfiguration time given by Eq. 6.16 as $\tau_{ij}^{(RIF)}$ and $\tau_{ij}^{(ZIF)}$ to differentiate between the reconfiguration times of RIF and ZIF, respectively.

It should be noted that the reconfiguration time defined here is related to the inter-monomer distance *vector* while the reconfiguration time in nsFCS studies is related to the fluctuations of the *absolute distance*. The choice of the former leads to a great mathematical simplification and affords simple analytic solutions. It may, however, be viewed as unphysical because, even if the distance between the monomers is not changing, the distance vector may relax through rotations, which, on the other hand, do not affect the FRET

signal. Fortunately, for Rouse or Zimm models, both the distance vector and absolute distance autocorrelation times are comparable and have the same scaling with chain length. Consequently, as previously shown in [106, 144], the prediction of Eq. 6.16 is in agreement with more realistic simulations of polymer dynamics. Of course, the inter-monomer distance autocorrelation time is equally easy to estimate numerically, if so desired, from the simulations described in Section 6.3.3.

6.4.2 Internal friction has additive effect on reconfiguration times

One of the most common methods [44] of quantifying internal friction in the conformational relaxation times of a polymer chain is to vary solvent viscosity, η_s, and to extrapolate the linear dependence to the absence of solvent viscosity (i.e., $\eta_s \to 0$). Internal friction then results in a finite positive intercept in the viscosity dependence. This approach has been used in the study of the conformational relaxation of myoglobin [150], the kinetics of protein folding [149, 151–154] and intrinsically disordered proteins and unfolded state proteins [144].

Applying this approach to RIF and using Eq. 6.8, Eq. 6.16 and Eq. 6.5, we find

$$\tau_{ij}^{(RIF)} = \tau_{ij}^{(Rouse)} + \tau_i \tag{6.17a}$$

Since $\tau_{ij}^{(Rouse)}$ is proportional to the solvent viscosity and assuming that τ_i is independent of the solvent viscosity, Eq. 6.17a predicts that all reconfiguration times will have exactly the same intercept, τ_i. In particular, for the

end-to end dynamics (i.e., $i = 0$ and $j = N$) and for long chains ($N \gg 1$), this approximately gives

$$\tau_{0N}^{(RIF)} \approx 0.8\tau_{Rouse} + \tau_i \tag{6.17b}$$

where $\tau_{Rouse} = \xi_s N^2 b^2/(3\pi^2 k_B T)$ is the relaxation time of the slowest Rouse mode. Thus, RIF predicts that the reconfiguration time exhibits a linear dependence on solvent viscosity with a finite intercept of τ_i (Fig. 6.1). In the limit of high internal friction ($\tau_i \gg \tau_{Rouse}$), the relaxation time of any mode is roughly τ_i (cf. Eq. 6.9). Consequently, relaxation of the distance vector between any two monomers will also be dominated by τ_i such that $\phi_{ij}(t) = \phi_{ij}(0)\exp(-t/\tau_i)$ and $\tau_{ij}^{(RIF)} \approx \tau_i$ and thus the reconfiguration time becomes independent of the monomer positions i and j. On the other hand, in the limit $\tau_i \ll \tau_{Rouse}$, the relaxation times of the Rouse model (Eq. 6.9) and the reconfiguration time of the Rouse model are respectively recovered [106]. Similarly, internal friction has an additive effect on the ZIF reconfiguration times, due to its similarity in mathematical structure to RIF. Incorporation of internal friction in the Zimm model similarly shifts the relaxation times of the polymer by an amount, τ_i (Eq. 6.13), and, likewise, each reconfiguration time $\tau_{ij}^{(Zimm)}$ by τ_i (Eq. 6.18). In particular, combining Eq. 6.12a, Eq. 6.16 and Eq. 6.5 yields the end-to-end reconfiguration time of the ZIF model, which for $N \gg 1$ approximately gives

$$\tau_{0N}^{(ZIF)} \approx 0.8\tau_{Zimm} + \tau_i \tag{6.18}$$

where $\tau_{Zimm} = \xi_s (N+1)^{3/2} b^2/(3\sqrt{3} k_B T \pi^{3/2})$ is the slowest relaxation time of the Zimm model and $\xi_s = 3\pi\eta_s b$. The factor of ~ 0.8 is coincidentally

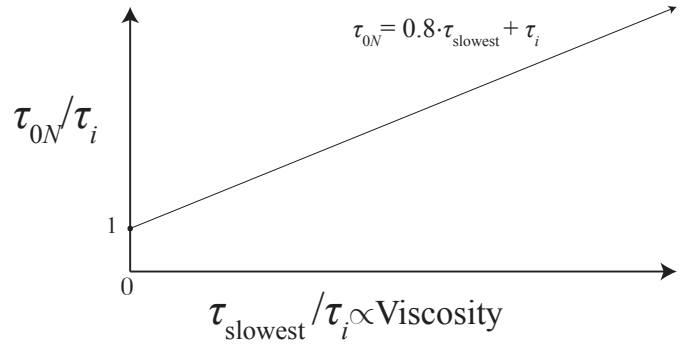


Figure 6.1: Both RIF (Eq. 6.17) and ZIF (Eq. 6.18) predict that the reconfiguration approximately exhibits a linear dependence on the slowest relaxation time of a free-draining and non-free draining chain, respectively, with a finite intercept of τ_i. In the absence of internal friction, the reconfiguration time is naturally smaller than the slowest relaxation time by a factor of ~ 0.8 due to the relaxation of the end-to-end vector through modes of higher frequency than the slowest mode.

the same as for the RIF model, despite softer decay in the mode dependence in Zimm ($p^{-3/2}$) as opposed to Rouse (p^{-2}). Similarly to RIF, ZIF predicts that the reconfiguration time exhibits a linear dependence on solvent viscosity with a finite intercept of τ_i (Fig. 6.1). The key difference between Eq. 6.17 and Eq. 6.18 stems from the difference in the chain length dependence of the relaxation times in the Rouse (N^2) and Zimm ($N^{3/2}$) models, respectively. This would generally lead to ZIF predicting a much faster reconfiguration time than RIF for very long chains ($N \gg 1$) while the quantitative difference between ZIF and RIF would be less significant for shorter chains. Similarly, both models would predict that internal friction would be more significant in short chains, which exhibit fast reconfiguration dynamics while exhibiting

negligible contributions of internal friction in the long chain limit ($N \gg 1$).

6.4.3 Comparison with experiment

An extension of RIF was used in a previous study to quantify the internal friction in the reconfiguration dynamics of unfolded proteins [144]. This extended model included a weak harmonic confinement potential to reproduce the experimentally observed unfolded state dimensions as well as the addition of beads to mimic the experimental dyes and linkers. Strictly speaking, the incorporation of these features complicated the simple mathematical structure of Eq. 6.17 such that internal friction was no longer additive. However, when comparing the reconfiguration times of the extended model with experimental data, deviations from additivity were minor and thus, for simplicity, the reconfiguration times were assumed to be additive. Incorporation of compaction and dyes and linkers is fairly straightforward in RIF and, thus, would also be straightforward in ZIF−for more details, see ref. [144].

6.5 Loop formation time of simple models

6.5.1 Introduction to the end-to-end loop formation problem

The second part of our study focuses on the end-to-end loop formation time, τ_{loop}, which is the average time that it takes for the ends of a polymer to diffusively come into contact with one another. Here, contact is defined as when the end-to-end distance, R, is less than a specified cutoff distance of R_c. The problem of end-to-end loop formation has been the focus of a number of

theoretical [3,13,20,23,25,28,31,66,92] and experimental [2,4,35] studies over the last few decades. In regards to the Rouse model, the loop formation time has been extensively studied [20,28] and is known to exhibit a chain length dependence of $\tau_{loop}^{(Rouse)} \propto N^2$ while being proportional with solvent viscosity, η_s. Likewise, the loop formation time of Zimm [20,66] is known to exhibit a weaker chain length dependence than Rouse (i.e., $\tau_{loop}^{(Zimm)} \propto N^{3/2}$) while also being proportionately to η_s.

In what follows, we examine the role of internal friction on the loop formation times of RIF and ZIF. Our analysis is motivated by loop formation measurements [101,159,160] as well as by the somewhat contentious issue of the effect of internal friction on the rate of protein folding [149]. While not analogous to folding, loop formation mimics an elementary folding step [2–5] and so our results will likely provide insights into the latter issue.

Two models describing the effects of internal friction on folding have been discussed in the literature. In one, internal friction has an additive effect, similar to Eqs. 6.17 and 6.18. If so, extrapolation of the linear solvent viscosity dependence to the limit of solvent absence (i.e., $\eta_s \to 0$) would result in a finite intercept that is independent of chain length and is equal to τ_i. In the other model, roughness of the energy landscape results in the multiplicative renormalization of the solvent viscosity η_s [145]. In the context of loop formation, this would imply that τ_{loop} exhibits the same chain length dependence as it does in the absence of internal friction and that loop formation time would approach zero in the limit of $\eta_s \to 0$, exhibiting no finite intercept.

Furthermore, it is possible that the contribution of internal friction is neither additive nor multiplicative, as a recent study by Netz and co-workers [157] suggests. Their study derived a passage time of incremental relaxation in the end-to-end distance (i.e., short time behavior) using RIF to demonstrate that internal friction is intrinsically entangled with solvent-related viscous drag.

In our study, we find from computer simulations that the contribution of internal friction is neither additive or multiplicative, and results in a nonlinear dependence on solvent viscosity with a finite loop formation time in the limit of solvent absence (Fig. 6.2). Furthermore, when internal friction is weak and solvent viscosity is the dominant source of friction, the loop formation time exhibits a linear dependence on viscosity with a finite intercept that is related to internal friction. However, due to the nonlinearity the loop formation time, the linearly extrapolated intercept grossly overestimates the actual loop formation time in the limit of solvent absence. These findings are qualitatively similar to the nonlinear viscosity dependence in the passage times reported by Netz and co-workers [157] that resulted from internal friction, despite our respective studies being based on very different considerations. To gain a better understanding of the contribution of internal friction in the loop formation time, we turn to carefully examining the asymptotic limits of high and low internal friction in the proceeding sections.

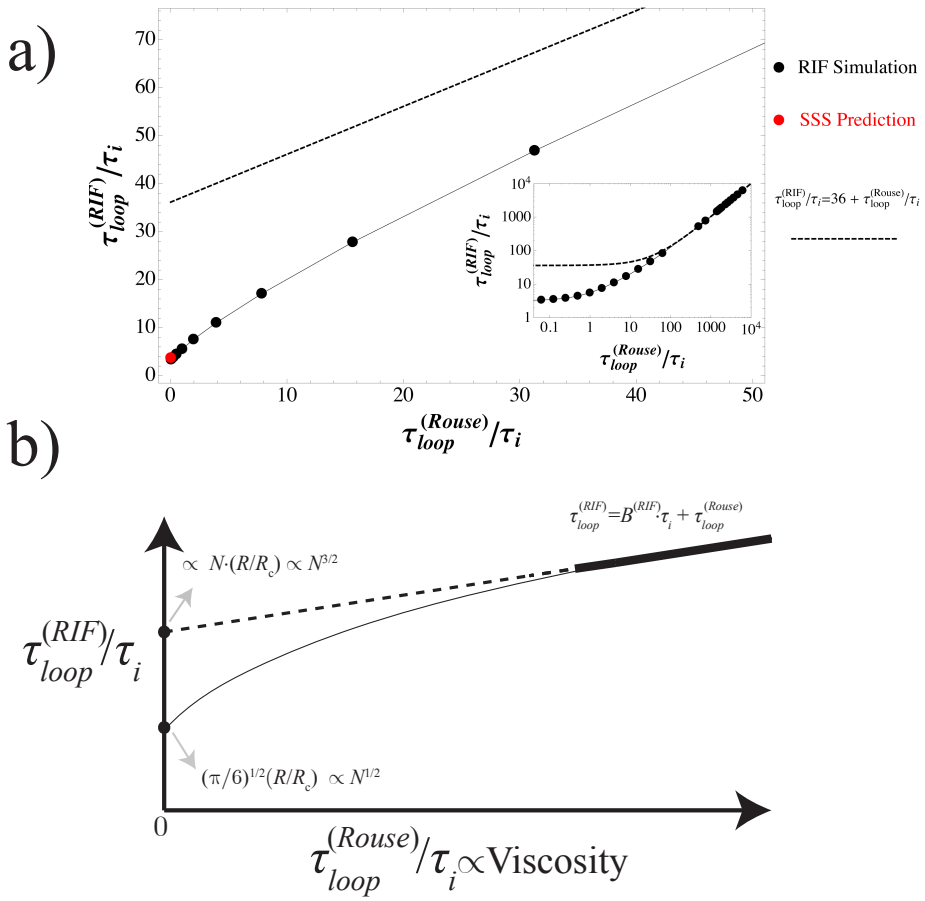


Figure 6.2: a) Here, the loop formation time is plotted as a function of the reciprocal of the internal friction timescale in dimensionless form for an RIF chain of length $N = 40$ with a capture radius of $R_c = 1.25$. Plotted alongside are the SSS prediction in the limit of $\tau_i \gg \tau_{Rouse}$ given by Eq. 6.26 and the linear fit of the simulation data in the limit of $\tau_i \ll \tau_{Rouse}$ (dashed line). The inset contains the log-log plot of the loop formation times over the entire range of internal friction explored by simulation. b) This diagram highlights our key findings for the role of internal friction in the loop formation time of RIF. Here, the loop formation data given by a thin, black line is bounded by the limiting expressions in the asymptotic limits of $\tau_i \gg \tau_{Rouse}$ and $\tau_i \ll \tau_{Rouse}$.

6.5.2 Loop formation: Limit of high internal friction

First, we seek to understand τ_{loop} in the limit of high internal friction, i.e., $\tau_i \gg \tau_{Rouse}$ and $\tau_i \gg \tau_{Zimm}$ for RIF and ZIF, respectively. In these respective limits, the RIF and ZIF modes become nearly degenerate such that Eq. 6.8 and Eq. 6.12a can both be re-written as

$$d\mathbf{X}_p/dt \approx -\tau_i^{-1}\mathbf{X}_p + \sigma_p \tag{6.19}$$

where σ_p is a delta-correlated Gaussianly distributed noise. To obtain a better understanding of the end-to-end loop formation dynamics in this limit, we first examine the dynamics of the end-to-end vector, $\mathbf{R}(t) = \mathbf{r}_N(t) - \mathbf{r}_0(t)$. Using Eq. 6.7 we can re-write $\mathbf{R}(t)$ as

$$\mathbf{R}(t) = \sum_{p=0}^{N} \mathbf{X}_p(t)(u_{pN} - u_{p0}) \tag{6.20}$$

Assuming that the end-to-end dynamics are overdamped (i.e., Brownian), Eq. 6.19 and Eq. 6.20 can be used to write the equation of motion governing $\mathbf{R}(t)$ as

$$d\mathbf{R}/dt = -\tau_i^{-1}\sum_{p=0}^{N} \mathbf{X}_p(u_{pN} - u_{p0}) + \sum_{p=0}^{N} \sigma_p(u_{pN} - u_{p0}) \tag{6.21a}$$

which simplifies to

$$d\mathbf{R}/dt = -\tau_i^{-1}\mathbf{R} + \mathbf{\Omega}(\mathbf{t}) \tag{6.21b}$$

Here, $\mathbf{\Omega}$ is delta-correlated Gaussianly distributed noise. In essence, the dynamics of the end-to-end vector in this limit is well described by a 3D overdamped harmonic oscillator with a relaxation time of τ_i. This effectively re-

places the polymer model with a dumbbell consisting of two end beads attached by a single spring, a model that has been studied and discussed by Doi [20, 161].

To understand τ_{loop} for the dumbbell model, we first turn to the approximations of Szabo-Schulten-Schulten [13] (SSS). In contrast to the end-to-end dynamics of a Rouse chain, the dynamics described by Eq. 6.21 are Markovian and, thus, an SSS-type or Kramers'-type approximation is adequate in this case. Provided that the capture radius R_c is much smaller than the rms end-to-end distance $\langle R^2 \rangle^{1/2}$, the SSS theory predicts the loop formation time to be given by

$$\tau_{loop} = \left(\frac{2\pi}{3} \right)^{3/2} \frac{\langle R^2 \rangle}{4\pi D R_c} \tag{6.22}$$

where D is the end-to-end diffusion coefficient. We estimate this coefficient using the Einstein-Stokes' formula,

$$D = k_B T / \gamma \tag{6.23}$$

where γ is the effective friction coefficient of end-to-end dynamics, which is related to the relaxation time τ_i and the oscillator spring coefficient k by

$$\tau_i = \gamma / k \tag{6.24}$$

The spring coefficient k can be estimated from the potential of mean force of the end-to-end vector, $\mathbf{R}$. Since the probability distribution of $\mathbf{R}$ is described by a Gaussian distribution, the potential of mean force is a harmonic potential with an effective spring coefficient of

$$k = 3 k_B T / \langle R^2 \rangle \tag{6.25}$$

Substituting Eqs. 6.23, 6.24, and 6.25 into Eq. 6.22, we find that

$$\tau_{loop} = \left(\frac{\pi}{6}\right)^{1/2} \tau_i \frac{\langle R^2 \rangle^{1/2}}{R_c} \tag{6.26}$$

Eq. 6.26 holds for both the RIF and ZIF models in the high internal friction limit and agrees well with simulation data (Fig. 6.3). Alternatively, the mean first passage time to form a looped conformation where $R \leq R_c$ can be found exactly by solving the 3D diffusion equation in a harmonic potential with an absorbing boundary at $R = R_c$. The solution to this problem was previously given [161], although there seems to be a minor numerical error in their final answer which has propagated through their subsequent papers. Their (corrected) solution is identical to the SSS formula (Eq. 6.26) and is shown to be exact in the limit of $R_c \ll \langle R^2 \rangle^{1/2}$, which is supported by the simulation data in Fig. 6.3. Consequently, we find that internal friction is neither additive nor multiplicative in the limit of high internal friction. Instead, we find that τ_{loop} exhibits a chain length dependence of $\tau_{loop} \propto N^{1/2}$ that is significantly weaker than the anticipated scaling of N^2 and $N^{3/2}$ for the Rouse and Zimm models, respectively. Furthermore, the SSS prediction for τ_{loop} given by Eq. 6.26 is exact for the dumbbell model.

6.5.3 Loop formation: Limit of small internal friction

In the limit where $\tau_{Rouse} \gg \tau_i$, the loop formation time $\tau_{loop}^{(RIF)}$ should, of course, approach the Rouse limit $\tau_{loop}^{(Rouse)}$. Imagine that τ_i is fixed while the solvent viscosity and, consequently, τ_{Rouse} is increased. This corresponds to

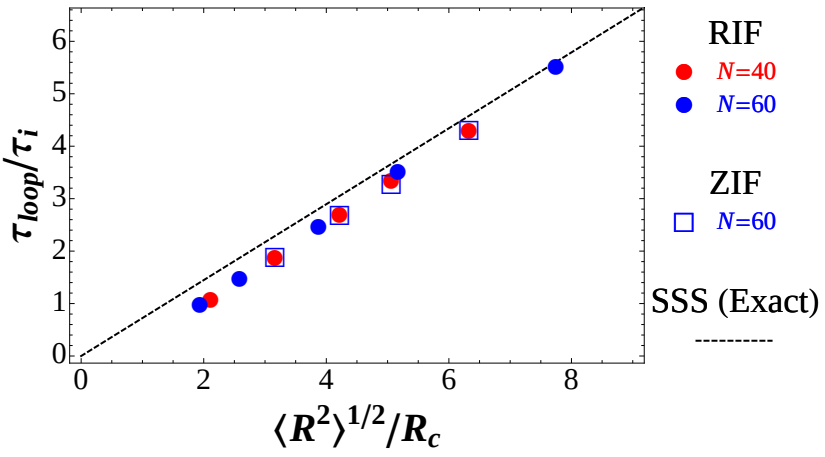


Figure 6.3: Here, the loop formation time predicted by SSS (Eq. 6.26) in the high internal friction limit is plotted as a function of the dimensionless parameter $\langle R^2\rangle^{1/2}/R_c$. Simulation data for both RIF and ZIF are plotted alongside the SSS result for comparison. All of the simulation data was found to collapse onto a universal curve, which agrees well with the SSS prediction. Furthermore, the agreement between simulation and SSS appears to improve as we approach the limit of $R_c \ll \langle R^2\rangle^{1/2}$, where the SSS prediction is known to be exact for the dumbbell model [161] with a characteristic relaxation time of τ_i.

the experimental scenario where the viscosity dependence of the loop formation time is measured. Just as in the case of nsFCS reconfiguration times, a nonzero intercept of $\tau_{loop}(\eta_s)$ extrapolated to zero viscosity (and, consequently, zero τ_{Rouse}) is expected to be the signature of internal friction. This intercept, however, is not equal to the limit of $\tau_{Rouse} \ll \tau_i$ estimated in the previous subsection because the dependence of $\tau_{loop}(\eta_s)$ or $\tau_{loop}(\tau_{Rouse})$ is not a straight line (Fig. 6.2). Rather, high-viscosity data (obtained at $\tau_{Rouse} \gg \tau_i$) extrapolated to zero viscosity give a higher value (Fig. 6.2). Using $\tau_{loop}^{(Rouse)}$ itself as

the measure of viscosity in this experiment, we write

$$\tau_{loop}^{(RIF)} = \tau_{loop}^{(Rouse)} + B^{(RIF)}\tau_i \tag{6.27a}$$

where we have surmised that the intercept $B^{(RIF)}\tau_i$ is proportional to τ_i. The dimensionless proportionality coefficient $B^{(RIF)}$ may, of course, depend on chain length and the capture radius. It is further expedient to divide by τ_i, which results in the dimensionless form of Eq. 6.27a:

$$\tau_{loop}^{(RIF)}/\tau_i = \tau_{loop}^{(Rouse)}/\tau_i + B^{(RIF)} \tag{6.27b}$$

Written in this manner, the dimensionless loop formation time in the weak internal friction limit exhibits a linear dependence on the dimensionless parameter $\tau_{loop}^{(Rouse)}/\tau_i$ with a slope that is, by construction, 1 and with a finite intercept of $B^{(RIF)}$. If this contribution of internal friction were purely additive, we would expect $B^{(RIF)}$ to simply be equal to 1 but, as will be seen below, this is not the case. Instead, this coefficient depends on both the capture radius R_c and the chain length N. Assuming that these dependences are power laws, we can rewrite them in a general form involving two dimensionless parameters, $\langle R^2\rangle^{1/2}/R_c$ and N:

$$B^{(RIF)}(N) \propto (\langle R^2\rangle^{1/2}/R_c)^{\delta} N^{\epsilon} \propto N^{\epsilon+\delta/2} \tag{6.28}$$

We have used computer simulations of RIF to independently determine both scaling exponents, δ and ϵ, by separately varying each dimensionless parameter. In doing so, we chose to vary $\tau_{loop}^{(Rouse)}/\tau_i$ by varying τ_i rather than the solvent viscosity, which is a more convenient yet equivalent approach.

Our data (Fig. 6.4) shows that δ is close to 1. Note that our numerical values are not strictly in the limit $b \ll R_c \ll \langle R^2 \rangle^{1/2}$, where Eq. 6.28 is expected to be universal, and so some finite size effects may be present. Likewise, we find that the scaling exponent ϵ is close to 1 (Fig. 6.4). Thus, we find that the dimensionless intercept of Eq. 6.27b is not simply equal to 1, as we would expect in the additive internal friction case, but rather exhibits a dependence on the capture radius of $B^{(RIF)} \propto R_c^{-1}$ and a total chain length dependence of $B^{(RIF)} \propto N\langle R^2 \rangle^{1/2}/R_c \propto N^{3/2}$ (noting that $\langle R^2 \rangle = Nb^2$ for our simple polymer). Notice that the chain length dependence of this intercept is weaker than the N^2 of the solvent controlled loop closure time but stronger than the $N^{1/2}$ dependence of the actual $\eta_s \to 0$ limit (Section 6.5.2), which is another manifestation of the nonlinearity of the viscosity dependence of the loop closure time. Similarly to the RIF case, the loop formation time of ZIF can be expressed as

$$\tau_{loop}^{(ZIF)}/\tau_i = \tau_{loop}^{(Zimm)}/\tau_i + B^{(ZIF)} \tag{6.29}$$

where

$$B^{(ZIF)}(N) \propto (\langle R^2 \rangle^{1/2}/R_c)^{\delta} N^{\epsilon} \tag{6.30}$$

We find from our data (Fig. 6.5) that δ is in the range $\delta \approx 1.15 - 1.38$ for $N = 60 - 120$ and ϵ is in the range $\epsilon \approx 0.63 - 0.92$ for $\langle R^2 \rangle^{1/2}/R_c \approx 4.22 - 7.91$. The variation in these indices is presumably due to finite-size effects. Indeed, the end-to-end loop formation times of our Zimm chain was found to exhibit a chain length dependence of $\propto N^{1.3}$ (data not shown) instead

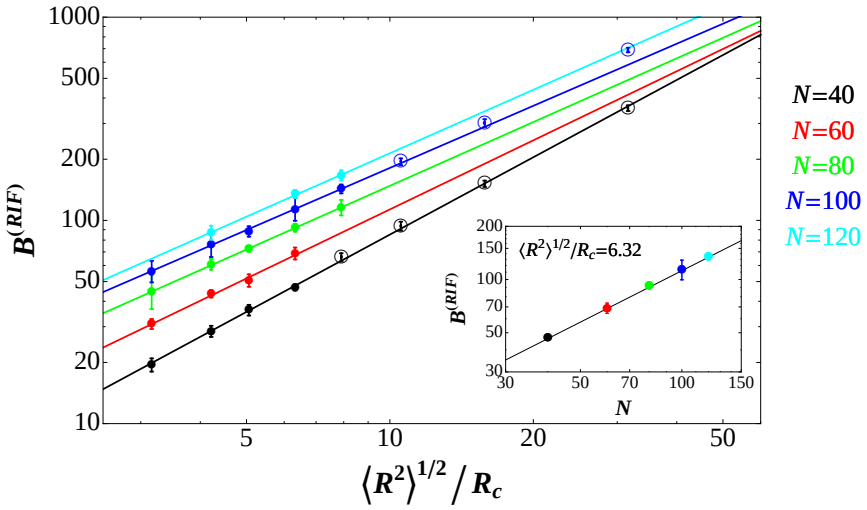


Figure 6.4: Simulations of RIF chains of different length N and capture radii R_c were used to explore the intercept of Eq. 6.28 in the limit of $\tau_i \ll \tau_{Rouse}$. For fixed N, $B^{(RIF)}$ is well described by power law fits of the form $B^{(RIF)} \propto (\langle R^2\rangle^{1/2}/R_c)^{\delta}$ where $\delta = 1.27 \pm 0.03$, 1.13 ± 0.06, 1.04 ± 0.01, 1.02 ± 0.02, and 1.04 ± 0.03 for N=40, 60, 80, 100, and 120, respectively. All fits were performed on data represented by filled circles, which denotes simulation data approaching the limit of $b < R_c \ll \langle R^2\rangle^{1/2}$. Open circles denoted simulation data where the capture radius is smaller than the Kuhn length (i.e., $R_c < b$), which were not included in the fits. Furthermore, for fixed $\langle R^2\rangle^{1/2}/R_c$, $B^{(RIF)}$ is well described by power law fits of the form $B^{(RIF)} \propto N^{\epsilon}$ where $\epsilon = 1.17 \pm 0.02$, 1.04 ± 0.03, 0.99 ± 0.06, 0.966 ± 0.007, and 0.90 ± 0.03 for $\langle R^2\rangle^{1/2}/R_c$ =3.16, 4.22, 5.06, 6.32, and 7.91, respectively. Shown in the inset plot is a representative chain length dependence of $B^{(RIF)}$ where $\langle R^2\rangle^{1/2}/R_c = 6.32$.

of the theoretically predicted $\propto N^{3/2}$ universal law expected in the limit of $N \to \infty$ [66]. Notwithstanding deviations from universality, the overall chain length dependence $B^{(ZIF)} \propto N^{\epsilon+\delta/2} = N^{1.2} - N^{1.6}$ is roughly the same as in the RIF case, while the capture radius dependence is $B^{(ZIF)} \propto R_c^{-1.15} - R_c^{-1.38}$.

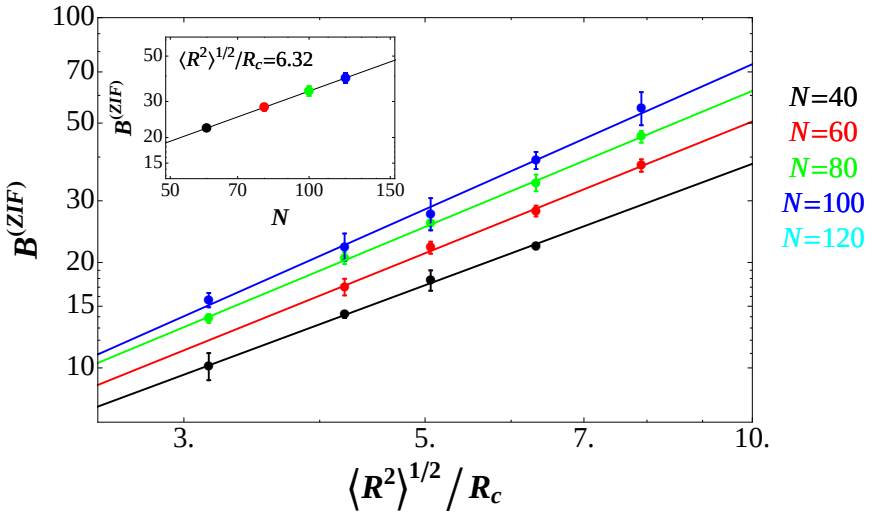


Figure 6.5: Simulations of ZIF chains of different length N and capture radii R_c were used to explore the intercept of Eq. 6.30 in the limit of $\tau_i \ll \tau_{Zimm}$. When N is fixed, $B^{(ZIF)}$ is well described by a power law fit of the form $B^{(ZIF)} \propto (\langle R^2\rangle^{1/2}/R_c)^{\delta}$ where $\delta = 1.15 \pm 0.03$, 1.25 ± 0.04, 1.29 ± 0.02, and 1.38 ± 0.05 for $N = 60$, 80, 100, and 120, respectively. Likewise, when $\langle R^2\rangle^{1/2}/R_c$ is fixed, $B^{(ZIF)}$ is well described by a power law fit of the form $B^{(ZIF)} \propto N^{\epsilon}$ where $\epsilon = 0.629 \pm 0.009$, 0.66 ± 0.05, 0.65 ± 0.06, 0.813 ± 0.004, and 0.92 ± 0.05 for $\langle R^2\rangle^{1/2}/R_c = 3.16$, 4.22, 5.06, 6.32, and 7.91, respectively. Shown in the inset plot is a representative chain length dependence of $B^{(ZIF)}$ where $\langle R^2\rangle^{1/2}/R_c = 6.32$.

6.5.4 Loop formation: Summary of viscosity and internal friction dependence

The viscosity and internal friction dependence of the loop formation time is summarized in Fig. 6.2 for the RIF case, although the general conclusions are nearly identical for the ZIF model. At very high viscosity, τ_{loop} exhibits a linear dependence on solvent viscosity. When extrapolated to zero viscosity, this dependence exhibits an intercept $B^{(RIF)}\tau_i$, which reflects inter-

nal friction. This intercept, however, is much larger than τ_i, in contrast to the additive internal friction model. Given that this intercept exhibits a weaker chain length dependence than that of τ_{loop} found at high viscosities, however, our result contradicts the multiplicative model, where internal friction would simply renormalize the solvent viscosity resulting in an overall N^2 scaling of the loop formation time.

In contrast to the linear growth of τ_{loop} with viscosity found at high values of viscosity, the actual viscosity dependence of $\tau_{loop}(\eta_s)$ is highly non-linear, resulting in a zero-viscosity limit, $\tau_{loop}(0)$, that is much smaller than the intercept $B^{(RIF)}\tau_i$ obtained via extrapolation of high-viscosity data. Instead, $\tau_{loop}(0)$ displays an even weaker chain length dependence of $N^{1/2}$ is inversely proportional to the capture radius, a result rigorously proven in Section 6.5.2.

6.5.5 Comparison with experiment

Next, we turn to comparing our simulated loop formation times with experimental loop formation times for unfolded peptides. A number of experimental studies have attempted to understand how changes in the dimensions of the unfolded state correspond to changes in the conformational dynamics[22, 37, 38]. In particular, Kiefhaber and co-workers [159, 160] examined unfolded poly(Gly-Ser) peptides consisting of 3 to 14 repeat units in an end-to-end loop formation study performed at different concentrations of denaturant. Their study found that decreasing the denaturant concentration corresponded to a decrease in both the chain dimensions and the effective end-to-end diffusion

coefficient [159], but also a decrease in the loop formation time [160]. Such a study provides an excellent opportunity for comparison with the simulated loop formation times of the CRIF model (See Section 6.7) due to the finding that greater compaction of the unfolded state correlated well with greater internal friction [144]. Here, the inclusion of compaction into our simple models such as RIF is critical when comparing to the experimental loop formation at low denaturant concentrations, where the spatial proximity of the polymer ends would be enhanced due to the compact conformation of the unfolded polymer.

When comparing our simulation data with the experimental loop formation times, we find remarkable agreement when internal friction is assumed to be absent in our model (Fig. 6.6) such that CRIF is simply reduced to a Rouse model with chain compaction. As noted by Kiefhaber and co-workers, the speed up in the loop formation times with decreasing denaturant concentration occurs due to the enhanced spatial proximity of the loop-forming groups afforded by chain compaction. Additionally, assuming the applicability of our simple models to describing unfolded proteins, the agreement between simulation and experiment would suggest that the contribution of internal friction is negligible in the peptides that were investigated. This conclusion is not a surprise considering that Kiefhaber and co-workers reported that all of their loop formation data was well described by near-linear power law dependences on solvent viscosity (i.e., a linear fit with an intercept of 0), a fact that was utilized in their viscosity correction of the loop formation times [160]. Although

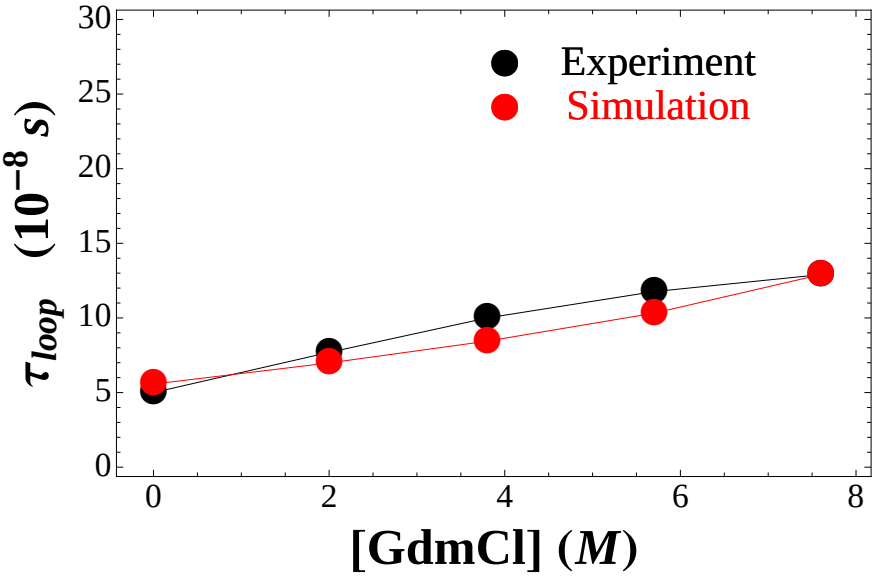


Figure 6.6: Here, the experimentally measured end-to-end loop formation times of a $(\text{Gly-Ser})_{14}$ peptide is plotted at different concentrations of denaturant [160]. Plotted in comparison are the simulated loop formation times of a Rouse chain with compaction, which is parameterized to reproduce the experimentally measured dimensions of the peptide at each denaturant concentration [159]. This agreement supports the assumption that the poly(Gly-Ser) peptide exhibits a negligible amount of internal friction. Here, our simulated polymer had a chain length of $N = 6$ where the Kuhn length was chosen to be equal to $b = 19$ Å(i.e., roughly 5 amino acids). A capture radius of $R_c \approx 3$ Åwas chosen such that the simulation data would match the experimental data at 7.6M GdmCl, where the experimental peptide was most expanded.

the agreement between experiment and simulation is reassuring, it highlights the need for experimental loop formation data for unfolded proteins and peptides with significant amounts of internal friction. Potential candidates for such a study would include unfolded peptide chains with bulky side groups that would potentially contribute to internal friction via steric hindrance.

6.6 Concluding Remarks

It has been postulated that a variety of mechanisms in proteins effectively give rise to a non-solvent related viscous drag (i.e., internal friction) that may significantly modulate the conformational dynamics of the protein molecule. Paying particular interest to unfolded proteins, we examine the role of internal friction in the conformational dynamics of simple random polymer models that phenomenologically incorporate internal friction as an adjustable free parameter−namely, Rouse with internal friction (RIF) and Zimm with internal friction (ZIF).

In the context of reconfiguration dynamics (i.e., FRET), we find that internal friction manifests itself as an additive contribution such that the reconfiguration times exhibit a linear dependence on solvent viscosity and a finite intercept equal to the relaxation time due to internal friction τ_i. Furthermore, due to the chain length dependence of Rouse and Zimm (i.e., N^2 and $N^{3/2}$, respectively), additivity in the reconfiguration times further suggests that internal friction would be negligible for very long polymer chains ($N \gg 1$) while exhibiting a stronger contribution in shorter chains.

In contrast, we find that the loop formation time τ_{loop} has a nonlinear solvent viscosity dependence and that internal friction does not simply result in an additive contribution or multiplicative contribution to τ_{loop}. The rather nontrivial dependence of τ_{loop} on various parameters defining the process thus offers a potentially critical experimental test of the RIF and ZIF models.

We have also have compared the simulated loop formation times of RIF with chain compaction with the experimental loop formation times of poly(Gly-Ser) peptides [159, 160], finding remarkable agreement between simulations and experiment when internal friction is taken to be absent in our simulations. This is consistent with the experimental findings [159, 160] suggesting that internal friction is negligible or absent in the poly(Gly-Ser) peptides that were explored. Furthermore, we find that accounting for the partial chain collapse is crucial when comparing to experimental data for chains that undergo a change in their dimensions, due to the enhanced spatial proximity of the loop-forming groups in a compact conformation that must be taken into account.

6.7 Appendix

6.7.1 Compacted Rouse with internal friction (CRIF)

When the amount of denaturant is decreased, the spatial dimensions of the unfolded polypeptide chain are typically found to decrease. While this effect can be captured by coil-globule transition theories [162], those theories deal with statistical properties rather than temporal evolution of the polymer. To account for the collapse of the chain without sacrificing the simplicity of the underlying model we adopt the approach described in ref. [144] and augment Eq. 6.4 with a potential that acts to pull each monomer towards the coordinate origin:

$$V_c = \frac{1}{2} k_c \sum_{i=0}^{N} r_i^2 \tag{6.31}$$

where k_c is the spring constant of this potential. We refer to this as a "compacted Rouse model with internal friction", or CRIF. As previously discussed [144], a potential of this form does not change the eigenmodes of the Rouse model and only modifies the equations of motion of the normal modes in Eq. 6.8 by shifting the spring constant of each mode to $k_p \to \kappa\lambda_p + k_c$. The rms distance between a pair of monomers, i and j, of the CRIF chain can be written as:

$$\langle (r_i - r_j)^2 \rangle_{\text{with compaction}} = 3k_B T \sum_{p=0}^{N} (k_c + \kappa\lambda_p)^{-1} (u_{pi} - u_{pj})^2 \qquad (6.32)$$

It should be noted that the decrease in the rms end-to-end distance of the chain could be mimicked in a simpler way, by reducing the Kuhn length b or, equivalently, increasing the bond spring constant κ. We feel, however, that this would be an inferior method because it would always preserve Gaussian chain statistics and would not account for the changes in the chain statistics upon its collapse. In contrast, CRIF was found to correctly reproduce the changes in the chain statistics, as compared with more realistic simulations of polymers undergoing the coil-globule transition [144].

When $k_c = 0$, Gaussian chain statistics are recovered. In our comparisons with experimental data we assume that this is the case at high denaturant concentrations. We then use k_c as an adjustable parameter at lower denaturant concentrations such that Eq. 6.32 matches experimentally observed values of the rms end-to-end distance.

Bibliography

[1] Jan Kubelka, James Hofrichter, and William A. Eaton. The protein folding 'speed limit'. *Current Opinion in Structural Biology*, 14(1):76–88, 2004.

[2] Oliver Bieri, Jakob Wirz, Bruno Hellrung, Mike Schutkowski, Mario Drewello, and Thomas Kiefhaber. The speed limit for protein folding measured by triplet-triplet energy transfer. *Proceedings of the National Academy of Sciences*, 96(17):9597–9601, 1999. 10.1073/pnas.96.17.9597.

[3] D. Thirumalai. Time scales for the formation of the most probable tertiary contacts in proteins with applications to cytochrome c. *J. Phys. Chem. B*, 103(4):608–610, 1999.

[4] S. J. Hagen, J. Hofrichter, A. Szabo, and W. A. Eaton. Diffusion-limited contact formation in unfolded cytochrome c: estimating the maximum rate of protein folding. *Proceedings of the National Academy of Sciences*, 93(21):11615–11617, 1996.

[5] D. E. Makarov and K. W. Plaxco. The topomer search model: A simple, quantitative theory of two-state protein folding kinetics. *Protein Science*, 12(1):17–26, 2003.

[6] X. J. Wang and W. M. Nau. Kinetics of end-to-end collision in short single-stranded nucleic acids. *J. Am. Chem. Soc.*, 126(3):808–813, 2003.

[7] F. Ricci, R. Y. Lai, and K. W. Plaxco. Linear, redox modified dna probes as electrochemical dna sensors. *Chem. Commun.*, 36:3768–3770, 2007.

[8] C. Fan, K. W. Plaxco, and A. J. Heeger. Electrochemical interrogation of conformational changes as a reagentless method for the sequence-specific detection of dna. *Proc. Natl. Acad. Sci. U.S.A.*, 100(16):9134–9137, 2003.

[9] B. Schuler and W. A. Eaton. Protein folding studied by single-molecule fret. *Current Opinion in Structural Biology*, 18(1):16–26, 2008.

[10] D. Nettels, I. V. Gopich, A. Hoffmann, and B. Schuler. Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc. Natl. Acad. Sci. U.S.A.*, 104(8):2655–2660, 2007.

[11] D. Nettels, A. Hoffmann, and B. Schuler. Unfolded protein and peptide dynamics investigated with single-molecule fret and correlation spectroscopy from picoseconds to seconds. *J. Phys. Chem. B*, 112(19):6137–6146, 2008.

[12] Ryan R. Cheng, Takanori Uzawa, Kevin W. Plaxco, and Dmitrii E. Makarov. The rate of intramolecular loop formation in dna and polypep-

tides: The absence of the diffusion-controlled limit and fractional power-law viscosity dependence. *The Journal of Physical Chemistry B*, 113(42):14026–14034, 2009.

[13] A. Szabo, K. Schulten, and Z. Schulten. First passage time approach to diffusion controlled reactions. *J. Chem. Phys.*, 72(8):4350–4357, 1980.

[14] Ryan R. Cheng and Dmitrii E. Makarov. End-to-surface reaction dynamics of a single surface-attached dna or polypeptide. *The Journal of Physical Chemistry B*, 114(9):3321–3329, 2010. PMID: 20151703.

[15] Ryan R. Cheng, Takanori Uzawa, Kevin W. Plaxco, and Dmitrii E. Makarov. Universality in the timescales of internal loop formation in unfolded proteins and single-stranded oligonucleotides. *Biophysical journal*, 99(12):3959–3968, 2010.

[16] Ryan R. Cheng and Dmitrii E. Makarov. Failure of one-dimensional smoluchowski diffusion models to describe the duration of conformational rearrangements in floppy, diffusive molecular systems: A case study of polymer cyclization. *The Journal of Chemical Physics*, 134(8):085104, 2011.

[17] A. V. Barzykin, K. Seki, and M. Tachiya. Diffusion-assisted long-range reaction between the ends of a polymer: Effective sink approximation. *J. Chem. Phys.*, 117(3):1377–1384, 2002.

[18] J. Z. Y. Chen, H.-K. Tsao, and Y.-J. Sheng. Diffusion-controlled first contact of the ends of a polymer: Crossover between two scaling regimes. *Phys. Rev. E*, 72(3):031804, 2005.

[19] P. Debnath and B. Cherayil. Dynamics of chain closure: Approximate treatment of nonlocal interactions. *J. Chem. Phys.*, 120(5):2482–2489, 2004.

[20] M. Doi. Diffusion-controlled reaction of polymers. *Chemical Physics*, 9(3):455–466, 1975.

[21] Friedman B. and O'Shaughnessy B. Theory of intramolecular reactions in polymeric liquids. *Macromolecules*, 26(18):4888–4898, 1993.

[22] L. J. Lapidus, P. J. Steinbach, W. A. Eaton, A. Szabo, and J. Hofrichter. Effects of chain stiffness on the dynamics of loop formation in polypeptides. appendix: testing a 1-dimensional diffusion model for peptide dynamics. *J. Phys. Chem. B*, 106(44):11628–11640, 2002.

[23] R. W. Pastor, R. Zwanzig, and A. Szabo. Diffusion limited first contact of the ends of a polymer: Comparison of theory with simulation. *J. Chem. Phys.*, 105(9):3878–3882, 1996.

[24] A. Podtelezhnikov and A. Vologodskii. Simulations of polymer cyclization by brownian dynamics. *Macromolecules*, 30(21):6668–6673, 1997.

[25] J. J. Portman. Non-gaussian dynamics from a simulation of a short peptide: Loop closure rates and effective diffusion coefficients. *J. Chem. Phys.*, 118(5):2381–2391, 2003.

[26] I. M. Sokolov. Cyclization of a polymer: First-passage problem for a non-markovian process. *Phys. Rev. Lett.*, 90(8):080601, 2003.

[27] S. Goundla, A. Yethiraj, and B. Bagchi. Nonexponentiality of time dependent survival probability and the fractional viscosity dependence of the rate in diffusion controlled reactions in a polymer chain. *J. Chem. Phys.*, 114(20):9170–9178, 2001.

[28] Ngo Minh Toan, Greg Morrison, Changbong Hyeon, and D. Thirumalai. Kinetics of loop formation in polymer chains. *The Journal of Physical Chemistry B*, 112(19):6094–6106, 2008.

[29] Z. S. Wang and D. E. Makarov. Rate of intramolecular contact formation in peptides: The loop length dependence. *J. Chem. Phys.*, 117(9):4591–4593, 2002.

[30] C. Yeung and B. Friedman. Cyclization of rouse chains at long- and short-time scales. *J. Chem. Phys.*, 122(21):214909, 2005.

[31] G. Wilemski and M. Fixman. Diffusion-controlled intrachain reactions of polymers. i. theory. *J. Chem. Phys.*, 60:866–877, 1974.

[32] Z. Guo and D. Thirumalai. Kinetics of protein folding: Nucleation mechanism, time scales, and pathways. *Biopolymers*, 36(1):83–102, 1995.

[33] D. E. Makarov, C. A. Keller, K. W. Plaxco, and H. Metiu. How the folding rate constant of simple, single-domain proteins depends on the number of native contacts. *Proc. Natl. Acad. Sci. U.S.A.*, 99(6):3535–3539, 2002.

[34] B. Fierz and T. Kiefhaber. End-to-end vs interior loop formation kinetics in unfolded polypeptide chains. *J. Am. Chem. Soc.*, 129(3):672–679, 2007.

[35] L. J. Lapidus, W. A. Eaton, and J. Hofrichter. Measuring the rate of intramolecular contact formation in polypeptides. *Proc. Natl. Acad. Sci. U.S.A.*, 97(13):7220–7225, 2000.

[36] K. Kawai, H. Yoshida, A. Sugimoto, M. Fujitsuka, and T. Majima. Kinetics of transient end-to-end contact of single-stranded dnas. *J. Am. Chem. Soc.*, 127(38):13232–13237, 2005.

[37] Takanori Uzawa, Ryan R. Cheng, Kevin J. Cash, Dmitrii E. Makarov, and Kevin W. Plaxco. The length and viscosity dependence of end-to-end collision rates in single-stranded dna. *Biophysical journal*, 97(1):205–210, 2009.

[38] J. J. Portman and P. G. Wolynes. Complementary variational approximations for intermittency and reaction dynamics in fluctuating environments. *J. Phys. Chem. A*, 103(49):10602–10610, 1999.

[39] J. Wang and P. G. Wolynes. Passage through fluctuating geometrical bottlenecks. the general gaussian fluctuating case. *Chem. Phys. Lett.*, 212(5):427–433, 1993.

[40] J. Wang and P. G. Wolynes. Survival paths for reaction dynamics in fluctuating environments. *Chem. Phys.*, 180(2-3):141–156, 1994.

[41] L. J. Lapidus, W. A. Eaton, and J. Hofrichter. Dynamics of intramolecular contact formation in polypeptides: Distance dependence of quenching rates in a room-temperature glass. *Phys. Rev. Lett.*, 87(25):258101, 2001.

[42] B. Bagchi, G. R. Fleming, and D. W. Oxtoby. Theory of electronic relaxation in solution in the absence of an activation barrier. *J. Chem. Phys.*, 78(12):7375–7385, 1983.

[43] P. G. De Gennes. *Scaling Concepts in Polymer Physics*. Cornell University Press, 1979.

[44] M. Doi and S. F. Edwards. *The Theory of Polymer Dynamics*. Clarendon Press/Oxford University Press, Oxford/New York, 1986.

[45] M. C. Murphy, I. Rasnik, W. Cheng, T. M. Lohman, and T. Ha. Probing single-stranded dna conformational flexibility using fluorescence spectroscopy. *Biophys. J.*, 86(4):2530–2537, 2004.

[46] C. Forrey and M. Muthukumar. Langevin dynamics simulations of genome packing in bacteriophage. *Biophys. J.*, 91(1):25–41, 2006.

[47] E. H. Yonemoto, G. B. Saupe, R. H. Schmehl, S. M. Hubig, R. L. Riley, B. L. Iverson, and T. E. Mallouk. Electron-transfer reactions of ruthenium trisbipyridyl-viologen donor-acceptor molecules: Comparison of the distance dependence of electron transfer-rates in the normal and marcus inverted regions. *J. Am. Chem. Soc.*, 116(11):4786–4795, 1994.

[48] A. Szabo and D. J. Bicout. Escape through a bottleneck undergoing non-markovian fluctuations. *J. Chem. Phys.*, 108(13):5491–5497, 1998.

[49] M. J. Heller. Dna microarray technology: Devices, systems, and applications. *Annu. Rev. Biomed. Eng.*, 4:129–153, 2002.

[50] I. Rasnik, S. A. McKinney, and T. Ha. Surfaces and orientations: much to fret about? *Acc. Chem. Res.*, 38(7):542–548, 2005.

[51] D. S. Talaga, W. L. Lau, H. Roder, J. Tang, Y. Jia, W. F. DeGrado, and R. M. Hochstrasser. Dynamics and folding of single two-stranded coiled-coil peptides studied by fluorescent energy transfer confocal microscopy. *Proc. Natl. Acad. Sci. U.S.A.*, 97(24):13021–13026, 2000.

[52] E. Eisenriegler, K. Kremer, and K. Binder. Adsorption of polymer chains at surfaces: Scaling and monte carlo analyses. *J. Chem. Phys.*, 77(12):6296–6320, 1982.

[53] J. Shang and E. Geva. A computational study of the correlations between structure and dynamics in free and surface-immobilized single polymer chains. *J. Phys. Chem. B*, 109(34):16340–16349, 2005.

[54] J. Feng, K. Y. Wong, G. C. Lynch, X. Gao, and B. M. Pettitt. Peptide conformations for a microarray surface-tethered epitope of the tumor suppressor p53. *J. Phys. Chem. B*, 111(49):13797–13806, 2007.

[55] K. Y. Wong and B. M. Pettitt. Orientation of dna on a surface from simulation. *Biopolymers*, 73(5):570–578, 2004.

[56] J. Shang and E. Geva. Computational study of a single surface-immobilized two-stranded coiled-coil polypeptide. *J. Phys. Chem. B*, 111(16):4178–4188, 2007.

[57] M. Friedel, A. Baumketner, and J.-E. Shea. Effects of surface tethering on protein folding mechanisms. *Proc. Natl. Acad. Sci. U.S.A.*, 103(22):8396, 2006.

[58] Z. Zhuang, A. I. Jewett, P. Soto, and J.-E. Shea. The effect of surface tethering on the folding of the src-sh3 protein domain. *Phys. Biol.*, 6(1):5004, 2009.

[59] Y. Zhang, A. Donev, T. Weisgraber, B. J. Alder, and J. J. Graham, M. D. De Pablo. Tethered dna dynamics in shear flow. *J. Chem. Phys.*, 130(23):234902–234902, 2009.

[60] T. A. Knotts IV, N. Rathore, and J. J. De Pablo. Structure and stability of a model three-helix-bundle protein on tailored surfaces. *Proteins: Structure, Function, and Bioinformatics*, 61(2):385–397, 2005.

[61] T. A. Knotts IV, N. Rathore, and J. J. de Pablo. An entropic perspective of protein stability on surfaces. *Biophys. J.*, 94(11):4473–4483, 2008.

[62] A. Anne and C. Demaille. Dynamics of electron transport by elastic bending of short dna duplexes. experimental study and quantitative modeling of the cyclic voltammetric behavior of 3‘-ferrocenyl dna end-grafted on gold. *J. Am. Chem. Soc.*, 128(2):542–557, 2006.

[63] A. Soranno, R. Longhi, T. Bellini, and M. Buscaglia. Kinetics of contact formation and end-to-end distance distributions of swollen disordered peptides. *Biophys. J.*, 96(4):1515–1528, 2009.

[64] A. Perico and M. Beggiato. Intramolecular diffusion-controlled reactions in polymers in the optimized rouse-zimm approach. 1. the effects of chain stiffness, reactive site positions and site numbers. *Macromolecules*, 23(3):797–803, 1990.

[65] A. Perico and C. Cuniberti. Dynamics of chain molecules. intramolecular diffusion controlled reactions for a pair of terminal reactive groups.

J. Polym. Sci.: Polym. Phys. Ed., 15(8):1435–1450, 1977.

[66] Barry Friedman and Ben O'Shaughnessy. Theory of polymer cyclization. *Physical Review A*, 40(10):5950–5959, 1989.

[67] B. Friedman and B. O'Shaughnessy. Short time behavior and universal relations in polymer cyclization. *J. Phys. II*, 1:471–486, 1991.

[68] J. L. G. Fernandez, A. Rey, J. J. Freire, and I. F. Pierola. Cyclization dynamics of flexible polymers. numerical results from brownian trajectories. *Macromolecules*, 23(7):2057–2061, 1990.

[69] M. Ortiz-Repiso, J. J. Freire, and A. Rey. Intramolecular reaction rates of flexible polymers. 1. simulation results and the classical theory. *Macromolecules*, 31(23):8356–8362, 1998.

[70] M. Ortiz-Repiso and A. Rey. Intramolecular reaction rates of flexible polymers. 2. comparison with the renormalization group theory. *Macromolecules*, 31(23):8363–8369, 1998.

[71] A. Rey and J. J. Freire. Numerical simulation of the cyclization dynamics for flexible chains with excluded volume. *Macromolecules*, 24(16):4673–4678, 1991.

[72] J.-H. Kim, W. Lee, J. Sung, and S. Lee. Excluded volume effects on the intrachain reaction kinetics. *J. Phys. Chem. B*, 112(19):6250–6258, 2008.

[73] E. A. DiMarzio. Proper accounting of conformations of a polymer near a surface. *J. Chem. Phys.*, 42:2101–2106, 1965.

[74] Lei Huang and Dmitrii E. Makarov. The rate constant of polymer reversal inside a pore. *The Journal of Chemical Physics*, 128(11):114903, 2008.

[75] T. Kawakatsu. *Statistical Physics of Polymers*. Springer-Verlag, Berlin/Heidelberg, 2004.

[76] J. Des Cloizeaux and G. Jannink. *Polymers in Solution: Their Modelling and Structure*. Clarendon Press, Oxford, 1990.

[77] B. Tinland, A. Pluen, J. Sturm, and G. Weill. Persistence length of single-stranded dna. *Macromolecules*, 30(19):5763–5765, 1997.

[78] Takanori Uzawa, Chiaki Nishimura, Shuji Akiyama, Koichiro Ishimori, Satoshi Takahashi, H. Jane Dyson, and Peter E. Wright. Hierarchical folding mechanism of apomyoglobin revealed by ultra-fast h/d exchange coupled with 2d nmr. *Proceedings of the National Academy of Sciences*, 105(37):13859–13864, 2008.

[79] Peter E Wright and H.Jane Dyson. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of Molecular Biology*, 293(2):321 – 331, 1999.

[80] Kevin W Plaxco, Kim T Simons, and David Baker. Contact order, transition state placement and the refolding rates of single domain proteins. *Journal of Molecular Biology*, 277(4):985 – 994, 1998.

[81] Kevin W. Plaxco, Kim T. Simons, Ingo Ruczinski, and David Baker. Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics. *Biochemistry*, 39(37):11177–11183, 2000.

[82] K. J. Oh, K. J. Cash, and K. W. Plaxco. Beyond molecular beacons: Optical sensors based on the binding-induced folding of proteins and polypeptides. *Chemistry: A European Journal*, 15(10):2244–2251, 2009.

[83] Arica A. Lubin and Kevin W. Plaxco. Folding-based electrochemical biosensors: The case for responsive nucleic acid architectures. *Accounts of Chemical Research*, 43(4):496–505, 2010. PMID: 20201486.

[84] Abhishek K. Jha, Andrés Colubri, Karl F. Freed, and Tobin R. Sosnick. Statistical coil model of the unfolded state: Resolving the reconciliation problem. *Proceedings of the National Academy of Sciences of the United States of America*, 102(37):13099–13104, 2005.

[85] David Shortle and Michael S. Ackerman. Persistence of native-like topology in a denatured protein in 8 m urea. *Science*, 293(5529):487–489, 2001.

[86] Michael Andrec, Anthony K. Felts, Emilio Gallicchio, and Ronald M. Levy. Protein folding pathways from replica exchange simulations and a kinetic network model. *Proceedings of the National Academy of Sciences of the United States of America*, 102(19):6801–6806, 2005.

[87] B. Zagrovic and V. S. Pande. Structural correspondence between the alpha-helix and the random-flight chain resolves how unfolded proteins can have native-like properties. *Nat. Struct. Biol.*, 10(11):955–961, 2003.

[88] Evan R. McCarney, James H. Werner, Summer L. Bernstein, Ingo Ruczinski, Dmitrii E. Makarov, Peter M. Goodwin, and Kevin W. Plaxco. Site-specific dimensions across a highly denatured protein; a single molecule study. *Journal of Molecular Biology*, 352(3):672 – 682, 2005.

[89] Ian S. Millett, Sebastian Doniach, and Kevin W. Plaxco. Toward a taxonomy of the denatured state: Small angle scattering studies of unfolded proteins. In George D. Rose, editor, *Unfolded Proteins*, volume 62 of *Advances in Protein Chemistry*, pages 241 – 262. Academic Press, 2002.

[90] C J Camacho and D Thirumalai. Theoretical predictions of folding pathways by using the proximity rule, with applications to bovine pancreatic trypsin inhibitor. *Proceedings of the National Academy of Sciences*, 92(5):1277–1281, 1995.

[91] Matthias J. Feige and Emanuele Paci. Rate of loop formation in peptides: A simulation study. *Journal of Molecular Biology*, 382(2):556 –

565, 2008.

[92] G. Wilemski and M. Fixman. Diffusion-controlled intrachain reactions of polymers. ii. results for a pair of terminal reactive groups. *J. Chem. Phys.*, 60:878–890, 1974.

[93] In-Chul Yeh and Gerhard Hummer. Peptide loop-closure kinetics from microsecond molecular dynamics simulations in explicit solvent. *Journal of the American Chemical Society*, 124(23):6563–6568, 2002.

[94] Isabella Daidone, Hannes Neuweiler, Sören Doose, Markus Sauer, and Jeremy C. Smith. Hydrogen-bond driven loop-closure kinetics in unfolded polypeptide chains. *PLoS Comput Biol*, 6(1):e1000645, 01 2010.

[95] Marco Buscaglia, Lisa J. Lapidus, William A. Eaton, and James Hofrichter. Effects of denaturants on the dynamics of loop formation in polypeptides. *Biophysical Journal*, 91(1):276 – 288, 2006.

[96] Peng Qu, Xinxing Yang, Xun Li, Xiaoxue Zhou, and Xin Sheng Zhao. Direct measurement of the rates and barriers on forward and reverse diffusions of intramolecular collision in overhang oligonucleotides. *The Journal of Physical Chemistry B*, 114(24):8235–8243, 2010.

[97] B. Friedman and B. O'Shaughnessy. Kinetics of intermolecular reactions in dilute polymer solutions and unentangled melts. *Macromolecules*, 26:4888, 1993.

[98] Yu-Jane Sheng, Pei-Hsien Hsu, Jeff Z. Y. Chen, and Heng-Kwong Tsao. Loop formation of a flexible polymer with two random reactive sites. *Macromolecules*, 37(24):9257–9263, 2004.

[99] Dana Doucet, Adrian Roitberg, and Stephen J. Hagen. Kinetics of internal-loop formation in polypeptide chains: A simulation study. *Biophysical Journal*, 92(7):2281 – 2289, 2007.

[100] Andreas Reiner, Peter Henklein, and Thomas Kiefhaber. An unlocking/relocking barrier in conformational fluctuations of villin headpiece subdomain. *Proceedings of the National Academy of Sciences*, 107(11):4955–4960, 2010.

[101] Steven A. Waldauer, Olgica Bakajin, and Lisa J. Lapidus. Extremely slow intramolecular diffusion in unfolded protein l. *Proceedings of the National Academy of Sciences*, 107(31):13713–13717, 2010. 10.1073/pnas.1005415107.

[102] Changbong Hyeon and D. Thirumalai. Kinetics of interior loop formation in semiflexible chains. *The Journal of Chemical Physics*, 124(10):104905, 2006.

[103] H. S. Chan and K. Dill. Intrachain loops in polymers: Effects of excluded volume. *J. Chem. Phys.*, 90(1):492–509, 1989.

[104] D.L. Ermak and J.A. McCammon. Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys.*, 69(4):1352–1360, 1978.

[105] J. Rotne and S. Prager. Variational treatment of hydrodynamic interaction in polymers. *J. Chem. Phys.*, 50:4831–4837, 1969.

[106] Dmitrii E. Makarov. Spatiotemporal correlations in denatured proteins: The dependence of fluorescence resonance energy transfer (fret)-derived protein reconfiguration times on the location of the fret probes. *The Journal of Chemical Physics*, 132:035104, 2010.

[107] Yoshitsugu Oono and Takao Ohta. The distribution function for internal distances in a self-avoiding polymer chain. *Physics Letters A*, 85:480 – 482, 1981.

[108] Beat Fierz, Andreas Reiner, and Thomas Kiefhaber. Local conformational dynamics in -helices measured by fast triplet transfer. *Proceedings of the National Academy of Sciences*, 106(4):1057–1062, 2009.

[109] Nicholas C. Fitzkee and George D. Rose. Reassessing random-coil statistics in unfolded proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 101(34):12497–12502, 2004.

[110] Zhisong Wang, Kevin W. Plaxco, and Dmitrii E. Makarov. Influence of local and residual structures on the scaling behavior and dimensions of unfolded proteins. *Biopolymers*, 86(4):321–328, 2007.

[111] H. S. Chung, J. M. Louis, and W. A. Eaton. Experimental determination of upper bound for transition path times in protein folding from

single-molecule photon-by-photon trajectories. *Proc. Natl. Acad. Sci. U.S.A.*, 106(29):11837–11844, 2009.

[112] T.-H. Lee, L. J. Lapidus, W. Zhao, K. J. Travers, D. Herschlag, and S. Chu. Measuring the folding transition time of single rna molecules. *Biophys. J.*, 92(9):3275–3283, 2007.

[113] J. Wang and P. G. Wolynes. Instantons and the fluctuating path description of reactions in complex environments. *J. Phys. Chem.*, 100(4):1129–1136, 1996.

[114] P. Faccioli, A. Lonardi, and H. Orland. Dominant reaction pathways in protein folding: A direct validation against molecular dynamics simulations. *J. Chem. Phys.*, 133(4):045104, 2010.

[115] P. Faccioli, M. Sega, F. Pederiva, and H. Orland. Dominant pathways in protein folding. *Phys. Rev. Lett.*, 97:108101, Sep 2006.

[116] W. E. Vanden-Eijnden and E. Vanden-Eijnden. Towards a theory of transition paths. *Journal of Statistical Physics*, 123(3):503–523, 2006.

[117] Philipp Metzner, Christof Schutte, and Eric Vanden-Eijnden. Illustration of transition path theory on a collection of simple examples. *The Journal of Chemical Physics*, 125(8):084110, 2006.

[118] E. Vanden-Eijnden. *Computer Simulations in Condensed Matter: From Materials to Chemical Biology.* Springer, 2006.

[119] C. F. Abrams and E. Vanden-Eijnden. Large-scale conformational sampling of proteins using temperature-accelerated molecular dynamics. *Proc. Natl. Acad. Sci. U.S.A.*, 107(11):4961–4966, 2010.

[120] W. E. Vanden-Eijnden and E. Vanden-Eijnden. Transition-path theory and path-finding algorithms for the study of rare events. *Annual Review of Physical Chemistry*, 61:391–420, 2010.

[121] Srabanti Chaudhury and Dmitrii E. Makarov. A harmonic transition state approximation for the duration of reactive events in complex molecular rearrangements. *The Journal of Chemical Physics*, 133(3):034118, 2010.

[122] Alexander M. Berezhkovskii, Gerhard Hummer, and Sergey M. Bezrukov. Identity of distributions of direct uphill and downhill translocation times for particles traversing membrane channels. *Phys. Rev. Lett.*, 97:020601, Jul 2006.

[123] Sergey V. Malinin and Vladimir Y. Chernyak. Transition times in the low-noise limit of stochastic dynamics. *The Journal of Chemical Physics*, 132(1):014504, 2010.

[124] Daniel M. Zuckerman and Thomas B. Woolf. Transition events in butane simulations: Similarities across models. *The Journal of Chemical Physics*, 116(6):2586–2591, 2002.

[125] Gerhard Hummer. From transition paths to transition states and rate coefficients. *The Journal of Chemical Physics*, 120(2):516–523, 2004.

[126] Bin W. Zhang, David Jasnow, and Daniel M. Zuckerman. Transition-event durations in one-dimensional activated processes. *The Journal of Chemical Physics*, 126(7):074504, 2007.

[127] Joseph D. Bryngelson, José Nelson Onuchic, Nicholas D. Socci, and Peter G. Wolynes. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins: Structure, Function, and Bioinformatics*, 21(3):167–195, 1995.

[128] D. J. Bicout and A. Szabo. Entropic barriers, transition states, funnels, and exponential protein folding kinetics: A simple model. *Protein Science*, 9(3):452–465, 2000.

[129] Elizabeth Rhoades, Mati Cohen, Benjamin Schuler, and Gilad Haran. Two-state folding observed in individual protein molecules. *Journal of the American Chemical Society*, 126(45):14686–14687, 2004.

[130] Jeffrey A. Hanson, Karl Duderstadt, Lucas P. Watkins, Sucharita Bhattacharyya, Jason Brokaw, Jhih-Wei Chu, and Haw Yang. Illuminating the mechanistic roles of enzyme conformational dynamics. *Proceedings of the National Academy of Sciences*, 104(46):18055–18060, 2007.

[131] Robert B. Best, Emanuele Paci, Gerhard Hummer, and Olga K. Dudko. Pulling direction as a reaction coordinate for the mechanical unfolding of

single molecules. *The Journal of Physical Chemistry B*, 112(19):5968–5976, 2008. PMID: 18251532.

[132] Olga K. Dudko, Gerhard Hummer, and Attila Szabo. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proceedings of the National Academy of Sciences*, 105(41):15755–15760, 2008.

[133] Olga K. Dudko, J. Mathe, Attila Szabo, Amit Meller, and Gerhard Hummer. Extracting kinetics from single-molecule force spectroscopy: Nanopore unzipping of dna hairpins. *Biophysical Journal*, 92(12):4188 – 4195, 2007.

[134] Irina V. Gopich and Attila Szabo. Single-macromolecule fluorescence resonance energy transfer and free-energy profiles. *The Journal of Physical Chemistry B*, 107(21):5058–5063, 2003.

[135] Elisha Haas, Ephraim Katchalski-Katzir, and Izchak Z. Steinberg. Brownian motion of the ends of oligopeptide chains in solution as estimated by energy transfer between the chain ends. *Biopolymers*, 17(1):11–31, 1978.

[136] E. Haas and I.Z. Steinberg. Intramolecular dynamics of chain molecules monitored by fluctuations in efficiency of excitation energy transfer. a theoretical study. *Biophysical Journal*, 46(4):429 – 437, 1984.

[137] Serdal Kirmizialtin, Lei Huang, and Dmitrii E. Makarov. Topography of the free-energy landscape probed via mechanical unfolding of proteins. *The Journal of Chemical Physics*, 122(23):234915, 2005.

[138] Pai-Chi Li and Dmitrii E. Makarov. Theoretical studies of the mechanical unfolding of the muscle protein titin: Bridging the time-scale gap between simulation and experiment. *The Journal of Chemical Physics*, 119(17):9260–9268, 2003.

[139] Dmitrii E. Makarov. Computer simulations and theory of protein translocation. *Accounts of Chemical Research*, 42(2):281–289, 2009.

[140] Z. S. Wang and D. E. Makarov. Nanosecond dynamics of single polypeptide molecules revealed by photoemission statistics of fluorescence resonance energy transfer: a theoretical study. *J. Phys. Chem. B*, 107(23):5617–5622, 2003.

[141] Peter Hanggi, Peter Talkner, and Michal Borkovec. Reaction-rate theory: fifty years after kramers. *Rev. Mod. Phys.*, 62:251–341, 1990.

[142] Yohichi Suzuki and Olga K. Dudko. Single-molecule rupture dynamics on multidimensional landscapes. *Phys. Rev. Lett.*, 104:048101, Jan 2010.

[143] R. Zwanzig. *Nonequilibrium Statistical Mechanics*. Oxford University Press, New York, 2001.

[144] Andrea Soranno, Brigitte Buchli, Daniel Nettels, Ryan R. Cheng, Sonja Muller-Spath, Shawn H. Pfeil, Armin Hoffmann, Everett A. Lipman, Dmitrii E. Makarov, and Benjamin Schuler. Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. *Proceedings of the National Academy of Sciences*, 2012. 10.1073/pnas.1117368109.

[145] R. Zwanzig. Diffusion in a rough potential. *Proceedings of the National Academy of Sciences*, 85(7):2029–2030, 1988.

[146] John J. Portman, Shoji Takada, and Peter G. Wolynes. Microscopic theory of protein folding rates. ii. local reaction coordinates and chain dynamics. *The Journal of Chemical Physics*, 114(11):5082–5096, 2001.

[147] Robert B. Best and Gerhard Hummer. Diffusive model of protein folding dynamics with kramers turnover in rate. *Physical Review Letters*, 96(22):228104, 2006. PRL.

[148] Aykut Erbas, Dominik Horinek, and Roland R. Netz. Viscous friction of hydrogen-bonded matter. *Journal of the American Chemical Society*, 134(1):623–630, 2011.

[149] Stephen J. Hagen. Solvent viscosity and friction in protein folding dynamics. *Current Protein and Peptide Science*, 11(5):1–11, 2010.

[150] A. Ansari, C. M. Jones, E. R. Henry, J. Hofrichter, and W. A. Eaton. The role of solvent viscosity in the dynamics of protein conformational

changes. *Science*, 256(5065):1796–1798, 1992. 10.1126/science.1615323.

[151] Troy Cellmer, Eric R. Henry, James Hofrichter, and William A. Eaton. Measuring internal friction of an ultrafast-folding protein. *Proceedings of the National Academy of Sciences*, 105(47):18320–18325, 2008. 10.1073/pnas.0806154105.

[152] Beth G. Wensley, Sarah Batey, Fleur A. C. Bone, Zheng Ming Chan, Nuala R. Tumelty, Annette Steward, Lee Gyan Kwa, Alessandro Borgia, and Jane Clarke. Experimental evidence for a frustrated energy landscape in a three-helix-bundle protein family. *Nature*, 463(7281):685–688, 2010. 10.1038/nature08743.

[153] Gouri S. Jas, William A. Eaton, and James Hofrichter. Effect of viscosity on the kinetics of alpha-helix and beta-hairpin formation. *The Journal of Physical Chemistry B*, 105(1):261–272, 2000.

[154] Linlin Qiu and Stephen J. Hagen. A limiting speed for protein folding at low solvent viscosity. *Journal of the American Chemical Society*, 126(11):3398–3399, 2004.

[155] Bhavin S. Khatri and Tom C. B. McLeish. Rouse model with internal friction: A coarse grained framework for single biopolymer dynamics. *Macromolecules*, 40(18):6770–6777, 2007.

[156] Bhavin S. Khatri, Masaru Kawakami, Katherine Byrne, D. Alastair Smith, and Tom C. B. McLeish. Entropy and barrier-controlled fluc-

tuations determine conformational viscoelasticity of single biomolecules. *Biophysical journal*, 92(6):1825–1835, 2007.

[157] Julius C. F. Schulz, Lennart Schmidt, Robert B. Best, Joachim Dzubiella, and Roland R. Netz. Peptide chain dynamics in light and heavy water: Zooming in on internal friction. *Journal of the American Chemical Society*, 134(14):6273–6279, 2012.

[158] Irina V. Gopich, Daniel Nettels, Benjamin Schuler, and Attila Szabo. Protein dynamics from single-molecule fluorescence intensity correlation functions. *The Journal of Chemical Physics*, 131(9), 2009.

[159] Andreas Moglich, Karin Joder, and Thomas Kiefhaber. End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation. *Proceedings of the National Academy of Sciences*, 103(33):12394–12399, 2006.

[160] Andreas Moglich, Florian Krieger, and Thomas Kiefhaber. Molecular basis for the effect of urea and guanidinium chloride on the dynamics of unfolded polypeptide chains. *Journal of Molecular Biology*, 345(1):153–162, 2005.

[161] Seiichi Sunagawa and Masao Doi. Theory of diffusion-controlled intrachain reactions of polymers. ii. *Polymer Journal*, 8(3):239–246, 1975.

[162] Guy Ziv, D. Thirumalai, and Gilad Haran. Collapse transition in proteins. *Physical Chemistry Chemical Physics*, 11(1):83–93, 2009.

Vita

Ryan Cheng was born in Taipei, Taiwan, the son of Nancy and Eric Cheng. At a young age, he immigrated to the United States of America with his family and became a naturalized citizen. In 2003, Ryan graduated from Marlboro High School in New Jersey and travelled to Pittsburgh to study at Carnegie Mellon University. From 2005-2007, he worked as an undergraduate researcher under the supervision of Professor Linda Peteanu and Dr. Gizelle Sherwood, who was a graduate student at the time. During the summer of 2006, Ryan was a participant in the Bioengineering and Bioinformatics Summer Institute (BBSI) at the University of Pittsburgh while mentored by Professor Ken Jordan. In 2007, Ryan received a B.S. in Chemistry from Carnegie Mellon with University Honors and Mellon College of Science Research Honors. After his graduation, Ryan moved to Austin, Texas to pursue a Ph.D. in Chemistry at the University of Texas at Austin under the supervision of Professor Dmitrii Makarov.

Permanent address: 142 Yorkshire Drive, Morganville, NJ 07751

This dissertation was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.